

73283

STIC-Biotech/ChemLib

From: Rao, Manjunath N.
Sent: Wednesday, August 14, 2002 12:47 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 10/003,759

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10C 01
Phone: 306-5681

Date: 8-14-02

Please search the following as soon as possible for application with serial number **10/003,759**

SEQ ID NO: 1, SEQ ID NO:3 and nucleotides 52-783 of SEQ ID NO:3 against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

SEQ ID NO: 2 and amino acids 40-261 of SEQ ID NO:2 against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Crystal Mall 1, USPTO.

Searcher: Manjunath Rao
Phone: 306-4501/4506
Location: Biotech Lab
Date Picked Up: 8/15/02
Date Completed: 8/20/02
Searcher Prep/Review: _____
Clerical: 10 min
Online time: 9 min

TYPE OF SEARCH:
NA Sequences: 3
AA Sequences: 2
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: ABSS02/03
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:50:56 ; Search time 6219.51 Seconds

(without alignments)
6315.467 Million cell updates/sec

Title: US-10-003-759-1

Sequence: 1 999cgagtcgaacgtcgggc.....tctccgacactacaacac 1877

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ot: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	1877	100.0	1877	6	AX339680	AX339680 Sequence
2	1837.8	97.9	1874	6	RM072637	U72637 Rhodothermus
3	786	41.9	786	6	AX339682	AX339682 Sequence
4	131.8	7.0	1116	6	AF233376	AF233376 Streptomy
5	131.8	7.0	1116	6	AR129926	AR129926 Sequence
6	131.8	7.0	1116	6	AR168360	AR168360 Sequence
7	105.8	5.6	1716	6	AR168362	AR168362 Sequence
8	105.2	5.6	1944	1	AF145250	AF145250 Rhodother
9	97.8	5.2	3527	1	SHU51222	U51222 Streptomyces
10	96.8	5.2	1470	1	SREGLS	X73953 S. rochei eg
11	89	4.7	1311	1	AF130408	AF130408 Streptomy
12	87.6	4.7	1565	1	SL004629	U04629 Streptomyces
13	87.6	4.7	1176	1	SCG11A	AL133210 Streptomy
14	69	3.7	11202	1	AE005123	AE005123 Halobacte
15	68.8	3.7	1262	1	AF335723	AF335723 Burkholder
16	68	3.6	125020	9	AF429315	AF429315 Homo sapi
17	66	3.5	14483	1	AE006992	AE006992 Mycobacte
18	66	3.5	67200	1	MTV017	AL021897 Mycobacte
19	63.2	3.4	125020	9	AF429315	AF429315 Homo sapi
20	57.6	3.1	1983	6	AX047403	AX047403 Sequence
21	56.6	3.0	45313	1	SCD95A	AL357432 Streptomy
22	56.6	3.0	201050	1	AL646064	AL646064 Ralstonia
23	55	2.9	36368	1	SC9B5	AL035206 Streptomy
24	54.8	2.9	346887	1	AP002995	AP002995 Mesorhizo
25	54.4	2.9	38995	1	SCF34	AL109974 Streptomy
26	54	2.9	34182	1	SC111	AL096849 Streptomy
27	53.6	2.9	77534	1	AF235504	AF235504 Streptomy
28	53	2.8	33779	1	SCGD3	AF096822 Streptomy
29	52.2	2.8	1849	1	SGSTRB1F	X78973 S. galbuis DS
30	52	2.8	6530	1	AB017438	AB017438 Streptomy
31	52	2.8	26500	1	SC9H1	AL356592 Streptomy
32	52	2.8	349116	1	AP003003	AP003003 Mesorhizo
33	51.4	2.7	3644	5	AF034576	AF034576 Gallus ga
34	51.4	2.7	10029	1	AF263011S2	AF263012 Streptomy
35	51.2	2.7	22115	1	SCC92	AL096838 Streptomy
36	51	2.7	77457	1	AF210249	AF210249 Streptomy
37	50.6	2.7	5525	1	SGSTRB1FG	X78974 S. glaucisce
38	50.6	2.7	14596	1	AE007168	AE007168 Mycobacte
39	50.6	2.7	25459	1	SGA6985	AJ006985 Streptomy
40	50.6	2.7	30302	1	MTGCG511	292774 Mycobacteri
41	50.6	2.7	40806	1	MSGY23	AD000016 Mycobacte
42	50	2.7	14939	1	AE005092	AE005092 Halobacte
43	50	2.7	29625	1	SCB22A	AL159178 Streptomy
44	50	2.7	36394	1	SCD35	AL160312 Streptomy
45	50	2.7	40104	1	SCF11	AL132662 Streptomy

ALIGNMENTS

RESULT 1	AX339680	1877 bp	DNA	linear	PAT 10-JAN-2002
LOCUS	AX339680	Sequence 1 from Patent WO0196382.			
DEFINITION	AX339680				
ACCESSION	AX339680.1	GI:18135682			
VERSION	AX339680.1				
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					

AX339680 Sequence
U72637 Rhodothermu
AX339682 Sequence
AF233376 Streptomy
AR129926 Sequence
AR168360 Sequence
AR168362 Sequence
AF145250 Rhodother
U51222 Streptomyces
X73953 S. rochei eg
AF130408 Streptomy
U04629 Streptomyce
AL133210 Streptomy
AE005123 Halobacte
AF335723 Burkholder
AF429315 Homo sapi
AE006992 Mycobacte
AL021897 Mycobacte
AF429315 Homo sapi
AX047403 Sequence
AL357432 Streptomy
AL646064 Ralstonia
AL035206 Streptomy
AP002995 Mesorhizo
AL109974 Streptomy
AL096849 Streptomy
AF235504 Streptomy
AF096822 Streptomy
X78973 S. galbuis DS
AB017438 Streptomy
AL356592 Streptomy
AP003003 Mesorhizo
AF034576 Gallus ga
AF263012 Streptomy
AL096838 Streptomy
AF210249 Streptomy
X78974 S. glaucisce
AE007168 Mycobacte
AJ006985 Streptomy
292774 Mycobacteri
AD000016 Mycobacte
AE005092 Halobacte
AL159178 Streptomy
AL160312 Streptomy
AL132662 Streptomy

/transl_table=11
/protein_id="CAD20684.1"
/db_xref="GI:18135683"
/translation="MNMAKAVLVSLLLFGCDMLFPDSDNGKEPEPEPVELCGR
KWARDVAGRYRVINNVGAETACILEVLETFNFTIADHDNGNNVAIPAIFGC
HMGACTSNGLPRRVOELSDVRSMTLPTRRNMAADIDMSPTNNAGNYSQAE
LMIMLMNGVMDGSGSRVATVELAGATMEVNYADMWNTIARRTPTTSVELDLKA
FIDDAVARGYRREPMTHAVETFEIEMEGAGLRSDFSVTQ"

BASE COUNT 331 a 582 c 632 g 332 t
ORIGIN

Query Match 100.0%; Score 1877; DB 6; Length 1877;
Best Local Similarity 100.0%; Pred. No. 3.6e-278;
Matches 1877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gggcagatcgaaagctggcgcgcccttcgagatcgaaagccttccacgctgaltcgc 60
DB 1 GGGCGAGTCGACAGCTCGCGCGCTCGCAGATCGAAGCGCTCTCCACGCTGATTCCG 60
QY 61 tggctggcagcagcagctggcgctggcgagcttcccttcaactcgtcagatctgcct 120
DB 61 TGGCTGGCGACGCACTGGCGCTGGCGAGATTTCCTTCACTTCCGACGATTGGCCCT 120
QY 121 tacgactacggcgacgacgcccgcacgacgctgctccgggaagcgcaagtgcgacccatg 180
DB 121 TACGACTACGGCGGCGGACGCCGACGACGCTGTCGGGAGAGCGAGTTGCCACCATG 180
QY 181 tggcgctgcttcaatagggatgctgctgacacacgagctgaggcgagacacacaccta 240
DB 181 TGGCGTGCCTCAATACGGGATGCTGCTGACACCGATGAGGGGAGACCAACACTA 240
QY 241 cctccggcgcgacacacgctgaagctggcgagcgcttgacatctggcgcttggcgaacgacta 300
DB 241 CCTCCGGCGGACACACTGGAAGTGGCGAGCGCTGCGCTTGGCGGCTGGCCCAACGACTA 300
QY 301 cggcttccggagatgctgcttaccatctacgaacgcaatggaacgtgagatgg 360
DB 301 CGGCTTCCGGGATGCTGCTTACCTCCGCTCCACTACGAACGCAATGGAACGTGAGATGG 360
QY 361 cggcctcgagctctatcttcgactatgccaagggagagctgcttgcggccggcgacgag 420
DB 361 CCGCTCGAGCTCTATTTCGCTATGCGGTAAGGAGGAGGACTGCTTGGCGCCGGCCGACGG 420
QY 421 cggcgcgagctcgtgacttgcgggaacggcgctgcttccacccggcgcggttgcggt 480
DB 421 CGCGCGGAGCTGCTGATTCGGGAGCCGGAACCGCTGCTTCCACCCGGCGGGGTACGGGT 480
QY 481 cgaagcggaacacgctgctgctgctgagctccgctcctccgactccgagcggtgcgcta 540
DB 481 CGAAGCGGAACGCTGCTGCTGCTGAGTCCGCGCGTCTCCGATCCGACGGCGGTGCGCTA 540
QY 541 tggcttggctccaatcagcagcagcagcgtcttcaacggcgcgccctgcgcgcttgcgc 600
DB 541 TGGCTGGTCCAAATACGGCGGACGCGGCTTCAACCGGGCGGCGCTTGCCTTGGCC 600
QY 601 ccttcgcagcagcagcagctgcggaagcgagctgagcgacgcaaacgggtgcttgcata 660
DB 601 CTTTCCGACCGACGACTGGCGGGAAGCGACTGAGCGACCAACCGGTGCTTGCATCGCA 660
QY 661 caaggagaccttgccttcaatgaatgagccttgccttgccttgccttgccttgccttgc 720
DB 661 CAGGGGCACTTCTGACTTGAAGTGCCTTGTGATTTCAATGGAATGAACGCTCAT 720
QY 721 gcgctggctacgctgcttgcgagccttgccttgccttgccttgccttgccttgccttgc 780
DB 721 GCGTGGGCTACTGCTGAGCGCTGATTGCTTGGATGGACAGGGCTCTTTCGCA 780
QY 781 tggcgacacggaaggaagcagcgagcctgagcccgagccgacgctgagctgtgagagcg 840
DB 781 TGGCGACACGGAAGGAAGCAGCGAGCTGAGCCGACCGACCGCTGACGCTGTGCGAGCG 840
QY 841 ctggagcgcgcgctgctgctggcgggcgctacccgggtgatacaaacgtatggggcg 900

DB 841 CTGGAGCGCGCGAGTGGCGCGGGGGCGGCTACCGGGTGATCAACAGTATGGGGCGC 900
QY 901 gggagacggcccgatgctatgtagtgcgacttggaagagggcaacttcagatacaacggcg 960
DB 901 GGGAGACGGCCCACTGATGAGTCCGACTGGAAACGGCAACTTACGATCAACGGGGC 960
QY 961 cga tcaacgaacgaagcaacacgtgacgagctatccgagcactactcgtgggtgcacactg 1020
DB 961 CGATCACGACAAAGGCAACACACTGGCGCCCTATCCGGGCATCTACTTGGGTGGCACACTG 1020
QY 1021 gggcgcttcaacgaagcaacttcggatctgcggcgcgctgtcagagagctgtccgagctgcg 1080
DB 1021 GGGCGCTCACAGCAAAATTCGGGATTCGGCGGGCGCTGAGAGACTGTCCGACGCTGGC 1080
QY 1081 caagagctgagcagctacgacgagctcagcgagcgcttgaaatgcccgtacagacactg 1140
DB 1081 CAGACCTGAGACCTACCCCGCATACAGAGGCGCCTGGAATGCCGCTTACGACACTG 1140
QY 1141 gttcag tcccgctcagcaatctccgcaacgctacagcggcgccgacgctgagtatctg 1200
DB 1141 GTTCACTCCCGTACGAAATTCGCGCAACGGCTACAGCGCGCGCGGCTGATGATCTG 1200
QY 1201 gctgaacctggaacgcgcgctgtagtgcggcgcgctgagcggctgtggcaacgctgagctg 1260
DB 1201 GCTGAACCTGGAAAGCGCGGCTGATGCCGGCGGCGACCCGCTGCGCACCTGGAACCTGGC 1260
QY 1261 cggcgcaaccttggagagctgtagtgcgacctgagaccttgaaatctacatcgctcactcgcg 1320
DB 1261 CGGGCGCACCTGGGAAGTCTGCTATGCCACTGTGAGACTGGAATTCATGCTGCTACCGGGC 1320
QY 1321 cagcagcccaacacacgctcggtagagcagcttgagccttgaaagccttca tcaacgaacggct 1380
DB 1321 CACGAGCGCCACCACTCGGTGAGCAGCTGACCTGGAAGCGCTTATCGACAGCCGCT 1380
QY 1381 cggcccgcgcttaca tccgcccggagctgtagtgcgcttgagctggagcgagggcttcgaact 1440
DB 1381 CGCCCGCGCTTACATCCGCGGAGTGTGATCTGCAATGCGGTGGAACGGGCTTCCAACTT 1440
QY 1441 ctggagggcgggcgctgctgctgcaagcgccga tcttccgtaacgctgtcag taaagctgt 1500
DB 1441 CTGGAGGGGGGGCGCGGCTGCGGAAGCGCGCAATTTTCGTAAGCGTACAGTACCTGCT 1500
QY 1501 cacaagggcaacacgctaaagcagaagaacacccgctcggggcgctatgcggcgcgct 1560
DB 1501 CACACGGGCGCACGCGTATGCGCCAGAGAACACCCGCTCGGGCGCTTATCGGGCGCGCT 1560
QY 1561 cgtcttgctcgtacttgcctgtagtgcgctacgagggcgctcagccggcgctgtgcttcc 1620
DB 1561 CGCTTGTGCTTACTTGTGATGCGCTTACGAGGGGCTCAGCGCGGTGCTGCTTCC 1620
QY 1621 tggagcgcttgcctgctgctgctgcaagcaacttaccgaaaagcaaggagagacctg 1680
DB 1621 TGGAGCGCTTCTGCTTCCGCGAGAGAACTTATCCGAAAAGCAAGGAGAGACTCG 1680
QY 1681 taaagcaagggcttgagcttgcagacaggaagcgaggttgtagtgaagcaacgaagcgct 1740
DB 1681 TAAAGCAAGGGCTTGTGCACTGCGACCAAGAGCAAGGAGTTGTGTAGCCACCAAGGCT 1740
QY 1741 ccgcaagatgagctcaccaccccgcaagcgcaacttcgctgctcctgcctgactacagcg 1800
DB 1741 CCGCGGAAGTGGGCTCACCGCGGACGCACTTGTGCTGCGCTCGCTGACTACAGAGG 1800
QY 1801 cgttggcgctgcttcaacggcgagcgtataaagctatacaggtcgaaatccggaacttcc 1860
DB 1801 CGTTGGCGCGTCTTCAAAACGCGAGGCTATTAAGCTATACAGCTGGAATCGCAACTTCT 1860
QY 1861 ccgacacttaacaaacac 1877
DB 1861 CCGACACTTACAAACAC 1877

RESULT 2

```

RMU72637 1874 bp DNA linear BCT 11-MAY-1999
DEFINITION Rhodothermus marinus cellulase (celA) gene, complete cds.
ACCESSION U72637
VERSION U72637.1 GI:2304960
KEYWORDS
SOURCE Rhodothermus marinus.
ORGANISM Rhodothermus marinus.
REFERENCE 1 (bases 1 to 1874)
AUTHORS Hallidorsdottir,S., Thorolfsson,E.T., Spilliaert,R.,
Heggvadsson,G.O., Kristjansson,J.K., Holst,O., and Eggertsson,G.
Cloning, sequencing and overexpression of a Rhodothermus marinus
gene encoding a thermostable cellulase of glycosyl hydrolase family
12
JOURNAL Appl. Microbiol. Biotechnol. 49 (3), 277-284 (1998)
MEDLINE 9824232
REFERENCE 2 (bases 1 to 1874)
AUTHORS Hallidorsdottir,S.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1996) Department of Molecular Biology,
Institution of Biology, Grensasvegur 12, Reykjavik 108, Iceland
FEATURES
source Location/Qualifiers
1..1874
/organism="Rhodothermus marinus"
/strain="T1378"
/db_xref="taxon:29549"
710..1493
/gene="celA"
710..1492
/gene="celA"
/EC_number="3.2.1.4"
/Note="belongs to family H of cellulase catalytic domains;
endo-1,4-beta-glucanase"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AAB65594.1"
/db_xref="GI:2304961"
/translation="MNVMRAVLVSLLLFLGCDMLFPPDGNKKEPEPEPEVEELGGR
MDARVAGGRVYINNVGAFRAHELDVPTTGRMNAAYIWMSPYNSNGISGAEF
HMARPARTRDCAARAGAVRAHELDVPTTGRMNAAYIWMSPYNSNGISGAEF
MILNMNGSVMPGSGRVAIVELAGATWEVYADWMDNTIATRTPTTSVELDKAF
IDDAVARGIRREWYLHAETGFEIWEAGAGLRADFSVTVO"
BASE COUNT 331 a 580 c 631 g 332 t
ORIGIN
Query Match 97.9%; Score 1837.8; DB 1; Length 1874;
Best Local Similarity 99.7%; Pred. No. 3,6e-272;
Matches 1872; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY 1 gggcgagtcgaagtcggcgcgccctcgagatcgaaagcgtcttccacgctgattcgc 60
Db 1 GGGCGAGTCGAAGTCGGCGCGCCCTCGCAGATCGAAGGCGCTTCCCGAGCTGATTGCG 60
QY 61 tgcgtgcgaacgacgtggcgctggtggcgagttcccttacttcgtlccagattgcctc 120
Db 61 TCGGTGCGACGACGTGGGCGCTGGGCGAGTTCCCTTACTTCGTCAGATTGCCCTT 120
QY 121 taagactaagggcggaagcccgcaagcagcgtgtccgggaagcgcaagtcgcaacatg 180
Db 121 TACGACTAGCGGCGCAAGCCCGCACGACGTCGCCGGAAGCGAGTTGCCACCATG 180
QY 181 tgcgcgtccctatacaggtatgttcgtgacacacgattgtggcgcaacacacata 240
Db 181 TGGCGCTGCCTCATACGGGATGATGCTGTACACACCATGTGGCGACACCAACACTA 240
QY 241 cctcgcgcgcgaacacgttgaagtgcgaagcgccttgcgttgcgttgcgcgaacta 300
Db 241 CCTCGCGCGCGACACGTAAGTGCGAGCGCTGCGGCTCTGGCGCTGCCAACGACTA 300

```

```

QY 301 cggcttcggygattgctgctactcgggtcccatctcagaaacgcatggaactgtaagatcg 360
Db 301 CGGCTTCGGYGATTGCTGCTACTCGGVTCCCATCTCAGAAACGCACTGGAACGTAGGATGG 360
QY 361 cggcctgcggtctatttcgcgtatgacgaagggaacgtgtcttcgcccgcgcgaagg 420
Db 361 CGGCCTGCGGTCTATTTCGCTATGACCGAAGGGGACGTGTCCTCGCCCGCCGACGG 420
QY 421 cggcgcgagatcgctgatttgcggagccgacgtgtcttcacccgcgcgcgttaagctg 480
Db 421 CGCGCGGAGATTGCTATTTCGGGACCGGACCGCTGCTTCCACCCGCGCGGTACGGGT 480
QY 481 cgaagcggaacgcctgtgtgtgtgagtcgcgcgtctccgaltccgcagcggtgccta 540
Db 481 CGAAGCGGAACGCTGTGTGTGTGAGTCGCGCGCTCTCCGATCCGACGGGCTGGCTA 540
QY 541 tgcgtgtccataaagccgcgaacgcacgctgttcaaccggcgccgcctccgcttcgc 600
Db 541 TGCCTGTCCAAATACGCCGACGCGACGCTGTCAACCGGCGGCTCGCCCTTCGCGC 600
QY 601 ctccgacccgacgacttggccggaagcgactgagcgaacgcggttctgcatgcga 660
Db 601 CTTCGCCACGACGACTGGCCGGAAGCGCACTGAGCAGCAACCGGCTTGCATGCCA 660
QY 661 caaggcgacttcgtactctgaagtgccttgcatttcaatggaataatgaacgtcat 720
Db 661 CAGGCGCACTTCTACTCTGAAGTGCCTTGTCAATTCATGGAATGAATGAACGTCA 720
QY 721 ggcgtgcgtactgttctgtgagcctgttatgtcgtttgtagtgcgttcttccga 780
Db 721 GCGTGCCTACTGTCTGTGAGCCTGTATTGCTGTTGATGTCGACTGTCTTTCCCA 780
QY 781 tggcgcaacgaaagaaacggaagcctgagcccgagccgacgctgagctgtgcgaag 840
Db 781 TGGCGCAACGGAAGAAAGAACCGGACGCTAGCCCGAGCGACCGCTGAGCTGTGCGGAG 840
QY 841 ctggagcgcgcgcgactgtgcgcggggcgctacgcggtgtatcaacaacgtatgggccc 900
Db 841 CTGGAGCGCGCGCGGATGTGGCCGGGGGCGCTACCGGCTATCAACACGTAAGGGGCGC 900
QY 901 ggaagccgcgcagtgatctgaagtcgagctggaagcggaacttcaacatacagaacg 960
Db 901 GGAGACCGCCCACTGATTCAGGTGCGACTCGAAGACGGCACTTACATCAGTACACGGG 960
QY 961 cgaatcaacgaacgaaacacgctggccgacttccggccatctacttcgggtgcacatt 1020
Db 961 CGATCAACGAAACGCAACACGTCGCGGCTATCCGGCATCTACTTCGGGCGCCACT- 1019
QY 1021 gggcgccctcgaagcaacttcggatttcgcgcgcgcgtgcgaagcgttcgaagctgcg 1080
Db 1021 GGGCGCCCTCGACGAGCAATTCGGGATTG- CGCGCGCCGTCGACGACCTGCCACGTGGC 1078
QY 1081 caagagctgagcgtcacgcgcgactcagacggcgcgctgaatctgcgcgttacaacatc 1140
Db 1081 CACGAGCTGGAGC-TCACGCCGATTCACGACGGCCCGCTGGAATGCCCTACGACATCTG 1137
QY 1141 gtccagtcgccgtacgaatctccggcaacggtctaacgcgcgcgcgtgcgttgcgt 1200
Db 1138 GTTCAGTCCCGTCACGAATTCGCGCAACGGCTACAGCGGCGCGGAGCTGATGATCTG 1197
QY 1201 gctgaactgaaacgcgcgcgtgtatgcgcgcgcgcgcgcgcgcgttgcacacgtgaa 1260
Db 1198 GCTGAACCTGAAACGGCGGCTGTATGCCGGGCGGACCCGCTGACCATCCGTGAATGTCG 1257
QY 1261 cgggcgcaccttggaagctgtgtatgcgaacttggaactgaatctacacacacgcgcg 1320
Db 1258 CGGGCGCACTTGGGAAGTCTGATATGCCACTTGGAATTAATGACCTGACCGCGG 1317
QY 1321 caagcgcacacacgctcggctgagtcgaagcttgaaacgttcaatcgaacgcgcgt 1380
Db 1318 CACGACGCCACACGCTCGGTGAGCGAGCTGACCTGAAGGCTTCATCAGACACGCGGT 1377
QY 1381 cgcgcgcgcgtacatccgcgcgagatgtgtatctgcacatgcgttggaacgcgcgt 1440

```



```

source      1. 1116
            /organism="Streptomyces sp. 11A68"
            /strain="11A68"
            /db_xref="taxon:133452"
gene         1. 1116
            /gene="cell12A"
            /gene="cell12A"
misc_feature 1. 93
            /gene="cell12A"
            /note="Region: signal sequence"
            1. 1116
            /gene="cell12A"
            /note="Region: catalytic domain"
            /gene="cell12A"
            /note="endo cellulase"
            /codon_start=1
            /transl_table=11
            /product="cellulase 12A"
            /protein_id="AAF91283.1"
            /db_xref="GI:9651813"
            /translation="MRSHPRSATMIVLVVLSGLLTPAAPAANOICDRYGTIT
            QDRYVONNRMTSATQICINVTGNGFELTDGSPVINGAPKSYPSVDGCHYMCAP
            RTLLPMRLISIGSAPSSVRYTNGVYNAADILADPTPTNGNRTPEIMVFNRG
            PVOPIGSPVIGAHVGRSMVEWTSNGSNDYISFLAPSAISMSFVDKPDQAVSHG
            LATPDMYLTISQAGPEPWGGTGLAVNSFSSAVNAGNGGPTGPACVOYSYHTW
            PGGFTVDITNTGSTPYDGMELDTPLPAGHTVMSVMALLISPAGAVTANSTSGNR
            IANGNOSFPGGSSGAGFTAPAGARLNGTSCTVR"
            94..756
            /gene="cell12A"
            /note="Region: catalytic domain"
            757..792
            /gene="cell12A"
            /note="Region: linker region"
            793..1113
            /gene="cell12A"
            /note="Region: cellulose binding domain"

BASE COUNT 191 a 432 c 340 g 153 t
ORIGIN
Query Match 7.0%; Score 131.8; DB 1; Length 1116;
Best Local Similarity 50.7%; Pred. No. 5.6e-11;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;
QY 807 ctgagcccgagccgacccgagctgtgagagcgctggagccgagctgtgagcgagg 866
DB 77 CGGCTCCCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 136
QY 867 ggcgtaccgagtgatgcaaacagatagtgagcgagcgagccgagctgagtgatcg 926
DB 137 ACCGGTAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 196
QY 927 gactggaacgagcgcaacttcacgatacagcgagcgagctgagcaaacgagcgatg 986
DB 197 CCGGCAACGGTTTCAGATCCACCAGCGAGCGAGGTTCCGATGCCAGCAACGGCG 256
QY 987 ccgactacgagccatctctctcctgagtgagcgagcgagcgagcgagcgagcgat 1046
DB 257 AATCTCATCTCCGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 316
QY 1047 tgcgcgagcgagtgagagagcgctgctcagagctgagcgagcgagcgagcgatga 1106
DB 317 TGCCCATGCGGATCAGCTGATCGGAGCGGCCCGCAGAGTGATCTCTACCGGTACAC 376
QY 1107 cgaacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1166
DB 377 GCAACGGGCTCTACAAACGCGCGGTAGACATCTGG-----CTGGACCGGACAC 430
QY 1167 acggtctacagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgatc 1226
DB 431 CCAACGGGCTGTAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 490
QY 1227 cggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgatg 1286
DB 491 CCATGGGTTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 550

```

```

QY 1287 ccgactgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1346
DB 551 GGAGCAACGGTTGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 610
QY 1347 agctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1406
DB 611 GCTTCAGCGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 670
QY 1407 ggtatcgtacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1466
DB 671 GGTACTCTACACGAGATCCAGCGGCGCTTCGAGCGCGGAGGAGGAGGAGGAGGAG 730
QY 1467 ggcgcgacttttcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1491
DB 731 TGAACCTGTTCTCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 755

RESULT 5
AR129926 AR129926 1116 bp DNA linear PAT 16-MAY-2001
LOCUS Sequence 2 from patent US 6187577.
DEFINITION
ACCESSION AR129926
VERSION AR129926.1 GI:14117823
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Jones,B.E., Van der Kleij,W.A.H., Van Solingen,P. and Weyler,W.
TITLE Cellulase producing Actinomyces cellulase produced therefrom and
method of producing same
JOURNAL Patent: US 6187577-A 2 13-FEB-2001;
FEATURES
source 1. 1116
location/Qualifiers
BASE COUNT 191 a 432 c 340 g 153 t
ORIGIN
Query Match 7.0%; Score 131.8; DB 6; Length 1116;
Best Local Similarity 50.7%; Pred. No. 5.6e-11;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;
QY 807 ctgagcccgagccgacccgagctgtgagagcgctggagccgagctgtgagcgagg 866
DB 77 CGGCTCCCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 136
QY 867 ggcgtaccgagtgatgcaaacagatagtgagcgagcgagccgagctgagtgatcg 926
DB 137 ACCGGTAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 196
QY 927 gactggaacgagcgcaacttcacgatacagcgagcgagctgagcaaacgagcgatg 986
DB 197 CCGGCAACGGTTTCAGATCCACCAGCGAGCGAGGTTCCGATGCCAGCAACGGCG 256
QY 987 ccgactacgagccatctctcctgagtgagcgagcgagcgagcgagcgagcgagcgat 1046
DB 257 AATCTCATCTCCGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 316
QY 1047 tgcgcgagcgagtgagagagcgctgctcagagctgagcgagcgagcgagcgagcgatga 1106
DB 317 TGCCCATGCGGATCAGCTGATCGGAGCGGCCCGCAGAGTGATCTCTACCGGTACAC 376
QY 1107 cgaacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1166
DB 377 GCAACGGGCTCTACAAACGCGCGGTAGACATCTGG-----CTGGACCGGACAC 430
QY 1167 acggtctacagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgatc 1226
DB 431 CCAACGGGCTGTAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 490
QY 1227 cggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgatg 1286
DB 491 CCATGGGTTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 550

```


QY	1469	tatctgcgtcggcggtggaacgagcgttcgaactctggaggcgcgagcggtctcgaagc	1468
Db	1125	TACCTACACACACATCCAGCGCGGCTTCGAGCGCTGGAGGCGGACCGCTTCGGCCGCTGG	1184
QY	1469	gccgattctccgtaacggtgca	1491
Db	1185	AACCTGCTTCCTCCGCGGTGCA	1207
RESULT	8	AF145250	1944 bp DNA linear BCT 09-JUN-1999
LOCUS	AF145250		
DEFINITION	Rhodothermus marinus putative O-methyltransferase gene, partial cds; and heat shock protein DnaJ (dnaJ) gene, complete cds.		
ACCESSION	AF145250		
VERSION	AF145250.1	GI:5020003	
KEYWORDS			
SOURCE	Rhodothermus marinus.		
ORGANISM	Bacteria: CFB group; Rhodothermus group; Rhodothermus.		
REFERENCE	1 (bases 1 to 1944)		
AUTHORS	Thorolfssdottir, E.T., Backman, V.M., Blondal, T., Kristjansdottir, S., Thordararnadottir, S.H., Palsson, A., Hauksdottir, H., Kristjansdottir, S., and Eggertsson, G.		
TITLE	Heat shock in Rhodothermus marinus: Cloning and sequence analysis of the groESL, dnaK and dnaJ genes		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 1944)		
TITLE	Backman, V., Thorolfssdottir, E.T., Kristjansdottir, S., Thordararnadottir, S.H. and Eggertsson, G.		
JOURNAL	Direct Submission		
AUTHORS	Submitted (23-APR-1999) Laboratory of Molecular Genetics, Institute of Biology, Grensasvegur 12, Reykjavik 108, Iceland		
TITLE	Location/Qualifiers		
FEATURES	1..1944		
SOURCE	/organism="Rhodothermus marinus"		
	/strain="ITI 376"		
	/db_xref="taxon:29549"		
	<1..343		
CDS	/codon_start=2		
	/transl_table=1		
	/product="putative O-methyltransferase"		
	/protein_id="A037972.1"		
	/db_xref="GI:5020004"		
	/translation="GRGRSDRRLGALDITLLEALDEGAAGTFDLAFIDAKASYDFA YEALQIVRPGGLIILDNILMGVRADPSNRREPSVEAIRLNKLKLDVALLLE LADGLGARRR"		
	478..1428		
gene	/gene="dnaJ"		
	478..1428		
CDS	/gene="dnaJ"		
	/codon_start=1		
	/transl_table=1		
	/product="heat shock protein DnaJ"		
	/protein_id="A037973.1"		
	/db_xref="GI:5020005"		
	/translation="MAQAVKDYEILGVPENATEEETIKAVRKLAREHPRNDPKPR AAEERFKEIOEAYSVLSDPEKRRQYDMARKNPFAGGCGPNCGRFVPTRECTVHPE TSGULEDLDEEERFGGLGDIFFSFFGEGCTGRHDPFAARRPRDVEAVQLPFEE ALRFGKDTITLPGQAKVIRINPKGVRFKTRLDGGCTTAGVPGANVYIILEVPHPR RFRREGQDITLTITINPLEMLGTTREIVGAVGRIRKIPPGTQPERLIRIGHGVBA TDTHTGDLVYEVQLEIRPNLSAEQRIILREAAKRAGLIRS"		
	1207..1428		
BASE COUNT	385 a 559 c 648 g 352 t		
ORIGIN			
Query Match	5.6%; Score 105.2; DB 1; Length 1944;		
Best Local Similarity	76.6%; Pred. No. 5.8e-07;		
Matches	167; Conservative 0; Mismatches 48; Indels 3; Gaps 3;		
QY	1648	agaactctccgaaagcaaggagagacactcgtgaacgacgagcgttctgcaactcgcacac	1707

[illegible]


```

/db_xref="GI:2209261"
/translation="MARGKRLVSLTAVFAATLGGIALTLFGOGNAOAHGVTMTPGSR
TYLMDAKTSTGSLDPTNPACAKALISGSAIKLNMFEVILDSMGAGACGVVPGKL
CSAGDRSPYNFTGVNABASDMRTHLTAGRTQYVHNSNNAAPSGFRYLSRPGSPS
TELGMDDLELITVTPNPOTGSPGTGIDGHTYNNLPSRSGDVAWFIOWHSDSEN
FEFSCDVDFDNGENGVETIRSGSGTDPDPTPTPTPTPTDPSHGCAVAVRYTWYS
GGFQGSVEVMHNHSTARDGMAYKMTPGAGAKSVWNGALTTGSDGAVVRSIDYNRS
IPPDGSVTFGFAITSTGNMLPYGSICVNP"

BASE COUNT      623 a 1270 c 1106 g 528 t
ORIGIN

Query Match      5.2%; Score 97.8; DB 1; Length 3527;
Best Local Similarity 48.5%; Pred. No. 6.9e-06;
Matches 304; Conservative 0; Mismatches 317; Indels 6; Gaps 1;

QY 865 ggggagctaccgggtgatcaacaagctatggggcgagagaccccgcatgtgaagt 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 747 gggccgctatgtgttcacagAACCGCTGGGGGGGAGTGGCCCACTGCTCAACCGC 806

QY 925 cggactgaaacgggcaacttcacgaatcaacggcgagatcaacgacaacgaagt 984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 807 TACCGACACGGGCTTCCGGGTGACCCAGGCGGAGCGCGCTCCACGACAGGTGCC 866

QY 985 ggcgcctatccggccatctacttcgggtgcacatggggcgccctgcagagaattcgag 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 867 GAAGTCTCTATCCCTCGGTCTTCAACGGCTGTCTACACCACTGCTGCCCGGAGCAA 926

QY 1045 atggcgagggcggtgcaggaagctgtgcagcgtgagcagagcttgagcgtcagccgat 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 927 CCTCCGGGCGGACAGGTACGGGTATCGCCAGCGCCCGACAGCATCTCTACGGCTTTGT 986

QY 1105 caacgacgggacggcttgatgacgcctacgacatctgttcaagtcaccgtacagaattccgg 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 987 CGGACACGCGCGGTATACAGCGCGGTACGACATCTGGCTGAGACCCACCAACCAAGAA 1046

QY 1165 caaagctacagggcgggcgccgagctgatatcttcgtcgaattgaaagcgcgctgata 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1047 CGGGGTGAACCC-----GTACCGAGATCATGATCTGGCTTCAACAGCTGCGCCGATCCA 1100

QY 1225 gccgagcgacgacggcggtgcacacgtgaaactgcccggggccacacttggaagtctcgata 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1101 GCCCATCGGCTCGCAGCGCGCGCACGCCCTCCGTGGCGGGCGGACCTGGCAGGTGTGGAG 1160

QY 1285 tgcgcgactgggaattgaatcagccacggcgacgaagccacacgctcgatgag 1344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1161 GGGCACAACAACGGCTCCACAGACGTATCTCTCTGTCGCCCTTTCGCCGTCACAGCTG 1220

QY 1345 cgagctggagccctgaaagccttcatacgacgacgggtccggcggtacatccggccgga 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1221 GAGCTTTCGACGTCATGACTTCTCTCCGGAACACCATTCGCCCGGCGCATGGCGCAAGAA 1280

QY 1405 gtgtatctgcacatgctgagagcgagcctcgaaactcgtggaagggcgagggcgagctcgag 1464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1281 CTGGTATCTTCACCAAGTGTCCAGGCAAGATTTCAGACCGGTGCGACAGACGGTCCGACTCCG 1340

QY 1465 aagcgccgaatttccgttaacggtgca 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1341 GGTGAACCTCTCTCTCTCAACCGTGAA 1367

RESULT 10
SREGS      1470 bp  DNA      linear  BCT 16-DEC-1994
LOCUS      S.rochei egIs gene.
DEFINITION X73953
ACCESSION  X73953
VERSION    GI:393391
KEYWORDS   cellulase; egIs gene; endoglucanase.
SOURCE     Streptomyces rochei.
ORGANISM   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomyces; Streptomyetaceae; Streptomyces.
REFERENCE  1 (bases 1 to 1470)

```

```

AUTHORS      Mastromei, G.
TITLE        Direct Submission
JOURNAL      Submitted (07-Jun-1993) G. Mastromei, Universita di Firenze, Dipt
            di Biologia Animale e Genetica, Via Romana, 17/19, 50125 Firenze,
            ITALY
REFERENCE    2 (bases 1 to 1470)
AUTHORS      Perito, B., Hanhart, E., Irdani, T., Iqbal, M., McCarthy, A.J. and
            Mastromei, G.
TITLE        Characterization and sequence analysis of a Streptomyces rochei A2
            endoglucanase-encoding gene
JOURNAL      Gene 148 (1), 119-124 (1994)
MEDLINE     95011642
FEATURES
    source          1..1470
                    /organism="Streptomyces rochei"
                    /strain="A2"
                    /db_xref="taxon:1928"
                    /clone="E.coli pCSF1"
                    90..103
                    /rpt_type=INVERTED
                    /rpt_type=INVERTED
                    112..125
                    /rpt_type=INVERTED
                    152..157
                    166..1314
                    /gene="egIs"
                    /gene="egIs"
                    166..1314
                    /gene="egIs"
                    /EC_number="3.2.1.4"
                    /codon_start=1
                    /transl_table=11
                    /product="cellulase"
                    /protein_id="CA52139.1"
                    /db_xref="GI:393392"
                    /db_xref="SPTREMBL:O59963"
                    /translation="MPRLRHPRITLRAVSALLTALALALATATAPADATICEE
                    RGSYVIGRVRVYONNRKGTSAQCVATADGSRVYQADGSPVTPNAPKSYSEVNGCH
                    YTNCSGPRALPARISGISAPSSISGCVDAVYNAASYDILDPDPRDGVNREIMI
                    WFNRYGLOPISQVGTASVAGRTVEVNSGCTNDYLSFAPASMSRSDVDNDFVR
                    ANVARGLOPISQVGTASVAGRTVEVNSGCTNDYLSFAPASMSRSDVDNDFVR
                    PAACTVSATNTNMPGFTANVTYNNGSAAPVDGMLATFLPSGQSVVHMANVASPSS
                    GAVTATGPAESRIRIAGGSQSFQGAAYSGSFAPAFQNLNCTACTV"
            stem_loop      1321..1340
            stem_loop      1351..1403
            stem_loop      1419..1453
BASE COUNT   211 a 579 c 475 g 205 t
ORIGIN

Query Match      5.2%; Score 96.8; DB 1; Length 1470;
Best Local Similarity 47.1%; Pred. No. 1.2e-05;
Matches 336; Conservative 0; Mismatches 372; Indels 6; Gaps 1;

QY 801 cggagcctgagcccgacggcgacgctgagctgtgcgagcgttgagacgcgcgcatgtgg 860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 254 CCGCGACCCCGCCCGACAGCGCGACACAGCATCTGCGAAGGTTCCGGCTCGACCGCTCA 313

QY 861 ccggggggcgctacccgggtgatcaacaacglatggggcgcgagacggcccgatgattcg 920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 314 TCCAGGGCGCGTATCGTTCAGAACACACCGGTGGGGCACACCGCCACCGCATGCTCA 373

QY 921 aggtcgactgaaacgggcaacttcagatcaacagggcgagatcaacgaaacggaaaca 980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 374 CCGCGACGACATCCCGGCTTCCGGGTACCGACGCGCGACGCGCTCGTCCGACCAACGGGCG 433

QY 981 acgtggcgctatccggcgcatctacttcgggtgcacactggggcgagcgtgcagagaatt 1040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 434 CGCGCAAGTCTGACCCGTGGGTCTTCAACGGCGCTGTACTACTACAGAACATTTTCCCGGCGCA 493

QY 1041 cgggaltgcgcggcgcgctgcagagagctgtccgagcgttgccagcagagctgagacgtcagc 1100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 494 CGCGGCTCCCGCGCGGAGTACAGGGGCAATCTCTCTCGCGCCACAGACATCTGTACGGGCT 553

QY 1101 cgatcagacggcgcgctggaatgacgctcgaacatctcggttcagtcgccgtaagaatt 1160

```


program of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.gov/jb/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid G11A lies on the AseI-6 genomic restriction fragment.

FEATURES

source

1. .41782

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid F37"

1. .1299

gene

/gene="xy1B"

misc_feature

/gene="xy1B"

/note="nominal overlap with Streptomyces coelicolor cosmid
2StG11"

CDS

<1. .1299

/gene="xy1B"

/note="SCG11A.01, xy1B, xy1ulose kinase, len: >432 aa;
highly similar to SW:XYLB_STRRU (EMBL:M73789) Streptomyces
rubiginosus xy1ulose kinase (EC 2.7.1.17) (xy1ulokinase)
xy1B, 481 aa; fasta scores: opt: 2553 z-score: 2795.9 E():
0; 89.2% identity in 427 aa overlap. Contains match to
pfam entry PF00370 FGcy, FGcy family of carbohydrate
kinases and two Prosite matches to entries PS00933 FGcy
family of carbohydrate kinases signature 2"

/codon_start=1

/transl_table=11

/product="xy1ulose kinase"

/protein_id="CAB61582.1"

/db_xref="GI:6468422"

/translation="DPRQMDALGELASQCEARAEAAVSGOQHGLVTLDARCP
VRPALMNDVRSAPQARRLDELGLGAKAMERTSSVSFTYTKRMALITHEEPAR
AVKAVRLPHDILTERLGEETTDGSDVSGTGMASGTEAVDETLARALDPAALPRV
VRPEVAGTVADGADLPESKGLTVAAGTGNAAALGILRPVMSLGTSGTAYAV
SORPADPTGTAGFADARCDMLPLACTLCTLAVDVYASLGLDREVEPTVTL
PLDGERPTNLPHSSGLLHGRHDTTGOILAAYDGAVSHLGCALDLVDADPSA
PLILGGARGTAGMOGTFRRLSGRPVOIPEARLVAIGAANAAGLLTGEDAAVARR
MNTAAGPYLDAVERDEATLNRITGVLSDAAPLIERDASR"

76. .1122

/gene="xy1B"

/note="Pfam match to entry PF00370 FGcy, FGcy family of
carbohydrate kinases, score 514.30, E-value 8.7e-151"

250. .288

/gene="xy1B"

/note="PS00933 FGcy family of carbohydrate kinases
signature 1"

874. .936

/gene="xy1B"

/note="PS00445 FGcy family of carbohydrate kinases
signature 2"

1519. .1523

1525. .2733

/gene="SCG11A.02"

1525. .2733

/gene="SCG11A.02"

/note="SCG11A.02, possible xylose repressor, len: 402 aa;
similar to SW:XYLR_BACSU (EMBL:M72248) Bacillus subtilis
xylose repressor xy1R, 384 aa; fasta scores: opt: 612
z-score: 642.4 E(): 2.1e-28; 29.4% identity in 381 aa
overlap. Contains match to pfam entry PF00480 ROK, ROK
family and a possible helix-turn-helix motif at residues
43. .64 (+3.15 SD)"

misc_feature

/gene="SCG11A.02"

/note="Pfam match to entry PF00480 ROK, ROK family, score
131.80, E-value 4e-38"

3041. .3646

/gene="SCG11A.03"

3041. .3646

/gene="SCG11A.03"

/note="SCG11A.03, probable amidase, len: 201 aa; similar
to SW:AMPD_ECOLI (EMBL:X15237) Escherichia coli
anhydro-N-acetylmuramyl-tripeptide amidase Ampd, 183aa;
fasta scores: opt: 211 z-score: 267.1 E(): 1.7e-07; 39.58
identity in 114 aa overlap and to Streptomyces coelicolor
SC2A11.21c; fasta scores: opt: 674 z-score: 751.6 E(): 0;
50.0% identity in 202 aa overlap. Contains match to Pfam
entry PF01510 Amidase_2, N-acetylmuramoyl-L-alanine
amidase"

/codon_start=1

/transl_table=11

/product="putative amidase"

/protein_id="CAB61584.1"

/db_xref="GI:6468424"

/translation="MERERTAPVPTRRLLKGAALATPYTLTSGTAAAGVADVP
SAEMLPASTNSYSSRRPTATPVYVYVHYQETADTSLFRPEQVSHYVRS
DGHAOCRETIDIAHAGNDYNTRSIGIEHGVADREIFNTAMYEQSAKLTAITC
AYGPKDRTTHIAHHEVPGSDHTDPCFWMTRYIRLVNFA"

3191. .3637

/gene="SCG11A.03"

/note="Pfam match to entry PF01510 Amidase_2,
N-acetylmuramoyl-L-alanine amidase, score 89.80, E-value
5.7e-23"

3811. .3815

3821. .5080

/gene="SCG11A.04"

3821. .5080

/gene="SCG11A.04"

/note="SCG11A.04, possible transcriptional regulator, len:
aa: similar to TR:O31511 (EMBL:Z99108) Bacillus subtilis
transcriptional regulator Acor, 605 aa; fasta scores: opt:
393 z-score: 446.5 E(): 1.7e-17; 34.7% identity in 199 aa
overlap"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator"

/protein_id="CAB61585.1"

/db_xref="GI:6468425"

/translation="MTDAMLALPEGADVERARALRRRAHETFEAGTVRRPRAVAD
SMRSVAGVPGDGTASVELMDGLGARAPRAHPRLSVVPLRELGFPAAGEHLLAV
CDAGGRLLMVGRHPRATRRACRMFPVPGARSESAVGNACPTAVAGRPVQAAEH
FIRVQPTCAAPVNDPRTRIGVLAVIDITGGDGLAPHSIGFYQAVARAESQDALI
TPERSASAEALRTALGRDELISADGRVRLRRHSELVILLAHPEGLTDELICAI
VEDETVPPVTLARELARLGLTGGRLASRPRLTMEVPSDTSVVERLRAGAVTAA
AAVAGPLPGSOAPAVGRLRRRLADGLAALACGPDLLADMAHAPGEGEDLDVWRA
LAAYRPATSSRLAALESELAAPNPW"

5205. .5209

5218. .6741

/gene="thca"

5218. .6741

/note="SCG11A.05, thca, aldehyde dehydrogenase, len:534

aa: highly similar to SW:THCA_RHOSN (EMBL:U17129)

Query Match 4.7%; Score 87.6; DB 1: Length 41782;
Best Local Similarity 47.5%; Pred. No. 0 00015;
Matches 298; Conservative 0; Mismatches 324; Indels 6; Gaps 1;

QY 865 gggcgctacacgggtatcaacaacglatyggcgcgagagagccacgtatgcatgagt 924
DB 27828 GGGCAGATGATGCTGTCAGAACACCGCTGGGGCTCCACCGCCACGACGCTGACGCGC 27887
QY 925 cggacgcgaaacggcgacactcacgatacaacggcgccgatacaagacagcgacaacgt 984
DB 27888 CACGACACCGGCTCCGGGTCACGACGCGCTGGGACCGCACCGACCGGGCGCC 27947
QY 985 ggcgcctatccgcacatctacttcgggtgacatggtggcgccctgacgagcaatcggg 1044
DB 27948 GAAGTGTATACCGCTGCTTCAACGGCTGCCACTACGAACTGTTACACGGCGCACGGC 28007
QY 1045 atgcgcgagcgctgcgagagctgctcgagctgacgacgagctggaacgctcagccgat 1104
DB 28008 CCTCCCGCTCCGGCTGACACCGCTTCGCGGCGCGCTCCAGCATCTCGACGCTTCGT 28067
QY 1105 caagcagcgccgctggaatgcgacatctggtacatcccgctacgataccgg 1164
DB 28068 CGACGGCGCTGTTACACGCGCTGTAACGACATATG6-----CTGACCGCGACGGCGCC 28121
QY 1165 caacgagctacagcgcgcgcgagctgatactgctgtaacttggaacgagcgcgctgat 1224
DB 28122 CACCGACGGGGGTGACACGACGAGATGATGATCTGTTCAACAGGGTGGCTCCGATCCA 28181
QY 1225 ggcggagcgagcgcgctgacacgctggaactgacgagcgccacactgggaagctcgta 1284
DB 28182 GCCCATCGCGCTCACCGGTGGGACGCGCTCCGTCGCGCGCGCGACCTGGAGCTGTGGAG 28241
QY 1285 tgcgcgactggagactggaatatacgcctacacggcgacgagcgccacacacgctcgtag 1344
DB 28242 CGCGGCGAACGCTGCAACACACTGCTGCTGCTGTCGTCGCGACCTCGGACGATCAGCGGCTG 28301
QY 1345 cgaagcgagactggaagcctcactacgacgagcgctgcgcgcgagctacatccgcgaga 1404
DB 28302 GACCTTGCAGATGATGAGACTGCTGCTCGGGGACGTCGCGCGGAGCTCGCGAAGAACGA 28361
QY 1405 gtgtatctgcatacgctgtagaagcgctcgaactctgggaagcgcgagctcgtag 1464
DB 28362 CTGGTACCTGACGACGCGCTTCAGCGGGGTTTCGAGCCGTGCGACACGCGCCGCGACTGGC 28421
QY 1465 aagcgcgacttttcgtaacggtgag 1492
DB 28422 CTGAACTCTCTCTCTCCACGCTCGAG 28449

RESULT 14
AE005123 11202 bp DNA linear BCT 12-FEB-2001
LOCUS Halobacterium sp. NRC-1 section 154 of 170 of the complete genome.
ACCESSION AE005123 AE004437
VERSION AE005123.1 GI:10581826
KEYWORDS
SOURCE Halobacterium sp. NRC-1.
ORGANISM Halobacterium sp. NRC-1.
Archaea; Euryarchaeota; Halobacteriales;
Halobacteriaceae; Halobacterium.
1 (bases 1 to 11202)
Ng,M.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sirogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
Dassarma,S.
TITLE From the cover: genome sequence of halobacterium species NRC-1

JOURNAL
PUBMED 11016950
REFERENCE 2 (bases 1 to 11202)
AUTHORS Ng,M.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,V., Sirogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
Dassarma,S.

TITLE Direct Submission
JOURNAL Submitted (14-JUL-2000) Institute for Systems Biology, 4225
Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES
source Location/Qualifiers
1..11202
/organism="Halobacterium sp. NRC-1"
/strain="NRC-1"
/db_xref="taxon:64091"
76..1281
/gene="meta"
/note="VNG2420C"
76..1281
/gene="meta"
/note="meta"
/codon_start=1
/transl_table=11
/product="homoserine O-acetyltransferase"
/protein_id="ANG20507.1"
/db_xref="GI:10581827"
/translation="MAALAVAVHNTVAKNYOYCTDGPMPTELDGFTFACGDTV
DSLTVAYRTGDTGTSNVLVCHALTGSOHVRGADGDRGMPDIYGPGRPIDTTEX
FVVCANVPGSCHGTDGPPADGCGPMCTAPFYPMTRCORLIDTLTGLVDRLHAY
VGGSVGNMALLDMARYPDDVRIAAVATAPLQOCGLNIAIARAIKGRNMNDV
YGGAHNPBGALAIROLQHVWLTSAWSADRGRTATPTPALPADPAEEFATPDE
STLDIADTFVAREFANSTLTLRLAMDITDLAAGPSYADALAAFDGALVLSFTGDM
HEPTAOSATVADAFRAGAAVHHVDSYGDHAFVDSAAVGPFLADFLAAGTGA
AVTDAITGCV"
1351..2646
/gene="hal"
/note="VNG2421G"
1351..2646
/gene="hal"
/note="hal"
/codon_start=1
/transl_table=11
/product="O-acetyl homoserine"
/protein_id="ANG20508.1"
/db_xref="GI:10581828"
/translation="MGHDGHLTNSVHAGRVDPATGARAPPLYQTSYAEADSADA
GOFALERDGYIYSRLMNPVETLQRLALLEGVAVATASGMALDIATFLARAGD
SVVAASDLXGGVTYLTLSAORRGVDTFVDVADADADATAVVLETVGNPS
LITPLELRIADIANDGVPLIADNTEFATPALATPDRHCAADIWSTTKWIGATGAG
CALVDAGSFDMDAHAADYPTAQENPATIHGCTFTDRRGDAFTYAIKRGRLDGNQ
SPEDAWQTLQKLETLPLRMQHCRNAQVAEHLRDPVNSVNTPLGLADHDHDAIT
YLDISYGMLTTFGEVDGYEAAOSVTEETTLASLANNVDAKTLVTHPASTHQDLTPE
AORAGCVPEWVRVSVGIEDPADIVADELFAIEAVGSA"
2715..5753
/gene="gld"
/note="VNG2422C"
2715..5753
/gene="gld"
/note="gld"
/codon_start=1
/transl_table=11
/product="glycolate oxidase subunit"
/protein_id="ANG20509.1"
/db_xref="GI:10581829"
/translation="MATRDLRSDDPSTDPADVDYTSVSRPGLVDDLREBVDGVR
FDSYRQLYATDASIEYALPVGRVPESTDDVAAVSVYCADRGIPVLRGGGCTIAG
AVNEAVVLDPRKAMNGFPGFPEARTAAOGCTTAAALNDELAAHGLKYAPDDPANGDK
SVLGAIGNNSGTAHSIQYERADGTLBSAEVVLADGVTYTTLGWDTDDLDLADPDGD

```
LEANIYADLRRLDEADRIEREPDLKRNVSQYNYDRLLADAEAGHANAARLLAGS
EGTIGITEAEVSLSEPPAETVAAMLTVDYDADADADAPVLDHDPASAEVMDVFLD
LAPDDEFADVYGLDPDGTDSLVEFYADSPAEARRVADLLADLPASATPADPAPG
ATTDDAPVAHVADALEAYDDDDROAHFMKRAKGLPTILSRTSDEKHYPIEDPAIDPDI
LPAFVADROAILLIDLDITVASYAHAGPCVLHVRLPYLTADENGAKKMAELIDRYTDLV
VEYGAVSGEHGDGRARQOMNRKTYGEDLMASFRTKAFDPDMLANPGVCGDAGE
IDMDHILKVAAPDDEFDAFDPALNMDTENGRGMETLCHGACGCTGHODTTGGVNCPT
YRAANEELSTRGRANMLRSAMNGLHPADPLDEFLTEVLDCVCGKCKLDCSPGVD
MAKLETEVKAHQADGVGLDRLEFGHTEALFHLGSLAPVSNMATLPGAGLAEKT
LGIAERDLPTFERRESLRAMFDARGSGRVEPAEAADHYTLAPDPYTNSTVAVGKAAV
RVLEAGVHYVALPGCVGSGRPAYSRGLIDRTRRIRARENVAAALAKVADGMDVYVEP
SDAYVVDYTHVLDLSSAPASVMAANTYVCETLDRHGIDDAANGFADGESLAIYHGHCH
OKAQRDHHAAGVLEKRAGYEVDPLDSTCCGNAGSGFYEAHEYMSQAIALIADQVAD
SDADITVAPGSSCRQSLSERDDVDTDGAPPHVOKLADAL"
complement(5763. .6470)
/gene="serB"
/note="VNG2423G"
complement(5763. .6470)
/gene="serB"
/note="serB"
/codon_start=1
/transl_table=1
/product="phosphoserine phosphatase"
/protein_id="AAG20510.1"
/db_xref="GI:10581830"
/transl_table="MLCWSAPRYCRMAGCYVAGRATPGMTLVAFEDDGTLAESMLDR
IAARAGVDEVAALITERAMKRELSTYADSLREAOIYAGLPESAANAAYDGYRLRNGAG
DLVAKLRDGGVAVVLTGCFKPGVNAADAGVADGVGNRLVNADELICAGVAGPL
VEGTRKDALRDACEAAGTTTPAAAVAVGDGANDVPMLDAAGTAIGVDPKRVADAHDT
VSSMDALGRVLDDHGLA"
6555. .8138
/gene="serA"
/note="VNG2424G"
6555. .8138
/gene="serA"
/note="serA"
/codon_start=1
/transl_table=1
/product="phosphoglycerate dehydrogenase"
/protein_id="AAG20511.1"
/db_xref="GI:10581831"
/transl_table="MRVLVTPDIADAGLRLRDAGHEVTTATDATGDALLDAVSDAAH
LIVBSGTAIVDAVFEAARDIVYARAGVGNIDDAATDHDHVMANAPAGVRAAE
HTVALAFAAARSIPDPAHRLDNGSAKDDYLGTELSGTKLGVGLRGVGEATVRLDLS
LGMDLVADPYIGEDRAQLGAELVDICTVARADFLIHVELTDETLGIGAEALAR
MDGCVIVNARGVDEEDALADADGVYAGALDLVFTTEPLPAASPLHDSITTP
HLGASTKAQENVAITDADQVVAALADGPVNALNAPSAERGAFFRIRPFVDLAETAG
TVAALFDPRIERVDVTYOGVEADEVDALVTAAQOGAPAGLEMOVNVNAPRAVEER
GIATVETNHRSEAYOSLVSYVNGDALETVSCTLFAFGDDPRILVEIDGFFVEAAPNG
HMLVARNHDPREVIGTIGCVLGTAEVNTAGMFAENAREARGGALTYINLDADVPARALD
ELAGDDRIVDVTSIELNGE"
8200. .8643
/gene="act"
/note="VNG2426G"
8200. .8643
/gene="act"
/note="act"
/codon_start=1
/transl_table=1
/product="acyl-CoA thioester hydrolase"
/protein_id="AAG20512.1"
/db_xref="GI:10581832"
/transl_table="MGSISDFEMENRENVDPDDTNINYASHGNNYKMMDEIGAMAAM
RHAGCTVATARINSLDFERPVPQGDICVIOAYATGHTSIVRLRARERDPRSGETE
PTTDSYFEVAVNDMDMPTAVVDLDVSGSERCELREALAGHDE"
8825. .9886
/gene="ald"
/note="VNG2429G"
8825. .9886
/gene="ald"
/note="ald"
/codon_start=1
/transl_table=1
```

```
/product="D-lactate dehydrogenase"
/protein_id="AAG20513.1"
/db_xref="GI:10581833"
/transl_table="MLSATPDADHITNTKAGPAFTTNGVIALLOYPSGRILSDLSRK
TLIVYGLMAAGGFVLTAFARSYALILVGAILVCGAALYPTPARALVADLXSTRQGR
AFGLHTAGSDJGSGAAGIAIVAVLALSLMRFAFVIVLITAVALLALHVSRESTEVE
FRSLAVRTPARLALISGTQFRMLIVATYATPWAIAFLLPFEDGEFAPVGTVA
FGALFVAVAIKAPPTAGTISDTISRRLVAGALVIGAAALSAAVVPSPVAAGVAVF
AAGLAFPPVQVSYLMDAFPESAGDGLAARVYIGLALGPYVAVSTPAAVNEIA
FMGLGALIVAVIARVY"
complement(9890. .11137)
/gene="thrc1"
/note="VNG2430G"
complement(9890. .11137)
/gene="thrc1"
/note="thrc1"
/codon_start=1
/transl_table=1
/product="threonine synthase"
/protein_id="AAG20514.1"
/db_xref="GI:10581834"
/transl_table="MTDLSLAGEFPACATDGVMLACVECGGLAFAPADVYRCPECSG
LLEARYADYPGFEARSGRQVRYADALPYDAGVSVQEGDPLYEAPDVAAAGVADVR
VKHEMNPTGSPKDRGKTVGVYARLGVLDLACSTSELSAALAAVYGRADPTVLTU
LPAGVVAAGKINQALHGARICEVDGNFDDCLDVYAEIARAGBATLILSLNPFRLGQ
KTIAPELILOSSTATGEMPRDRLVLPVNGANFALYKAFRELVAAGATDDEMPALTG
VOAGAAPMEVSEAVENRTERMDVETRAAIRIGNVNPAPKALPGVETGTAIVAV
SDDRITTDQMLMAGDGIQVPEPASAASVAGLLALARREGEISADERVCLTTHLLKDP
AANAAGGDTTPPADADVLDAL"
BASE COUNT 1670 a 4414 c 3626 g 1492 t
ORIGIN
```

Query Match 3.7%; Score 69; DB 1; Length 11202;
Best local Similarity 43.8%; Pred. No. 0.14;
Matches 300; Conservative 0; Mismatches 385; Indels 0; Gaps 0;

QY	778	cgatgcgacaacggaagaaacggagccctgaagcccgagccgctcgagctgtgag	837
DB	7058	CGACCCCTCATCGGAGGAGACCGGCGGCGGCGGCGGAGCTGTCGACATCGA	7117
QY	838	acgctggagcgcgcgagatgctggcgggggcgctccgggtgatcaaacgtatggg	897
DB	7118	GACCTCGTCGCGGCGGCGGCGGATTCCTCACGATCCAGTCCACTACCGACAGACGA	7177
QY	898	cgcgagacgcgcacatgcatitgaggttcggactgtgaaacgggcaacttcagatcacg	957
DB	7178	CGGCTCATCGGAGAGCGGGAAGTCCCGCGGAGAGGCGGCTACGTATCAAGTCGC	7237
QY	958	ggcgatcacgacaacgcaaacgctggcgcgctatccggccatctactcgggtgcca	1017
DB	7238	CCGCGCGGCGCTGTCGAGAGAGACCGCTGCGCGCGCGCCACAGACGCGTCAATCC	7297
QY	1018	ctggggcgctcgacagagaaatccgagattgcggggcggtgcaggaagcttccagct	1077
DB	7298	GGGGCGCGCCCTGACGCTGTCGACAGAGCCGCTCCGCGCGCTCCCGCTCACGA	7357
QY	1078	ggcgacgacgtcgacgctcagcgcgaatcagagcgggcgctggaatgcgcctacgacat	1137
DB	7358	CGCGGACAGCATCATCACACGCGCACACTCGGCGGCTGCACAGAGCGCGCCAGAGAA	7417
QY	1138	ctggttcagtlcccgltcagaaatccggcaacggcgtacagcgcgcgcgagctgatgt	1197
DB	7418	CGTCGCGCACCGACCGCGGACAGCTGCTGCGCGCGCTGCGCGGACCGGCTGTGAA	7477
QY	1198	ctggctgaactctgaaagcggtgtagtgcggcgcgcaacgscgtgtgccaacgttgaact	1257
DB	7478	CGCGCTGAACGGCCGACGCGGAGCGGTGCTGCAACCGCATCGCCCTGCTGCA	7537
QY	1258	ggcgagggccacacttggaagctctgtatgcgacacttggaactatcatcgccctaccg	1317
DB	7538	CGTCGGGAGACCGCGCGGACCGTCCCGCGGCGCTGTCGACCCCGGATGAGAGCGGT	7597

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:55:01 ; Search time 544 Seconds

(without alignments)
5923.986 Million cell updates/sec

Title: US-10-003-759-1

Perfect score: 1877

Sequence: 1 gggcgagtcgaacgtcgggc.....ttccgacacttaaacac 1877

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

3472872

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_032802:*

1: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.8	7.0	1116	AAV72359	Actinomyces sp. 3
2	131.8	7.0	1116	AAZ57029	Actinomyces cell
3	131.6	7.0	1715	AAZ57031	DNA sequence of an
4	57.6	3.1	1983	AAZ54513	Fructan exohydrolase
5	53.6	2.9	77536	AAAI4651	Nucleotide sequenc
6	51	2.7	58857	AAAS8471	Nucleotide sequenc
7	49.8	2.7	756	AAAI6766	Collagen-like poly
8	49.4	2.6	1404	AAAI8763	Wheat 1-deoxy-D-xy
9	49	2.6	756	AAO43032	Collagen-like poly

10	48.8	2.6	985	19	AAV64548
11	48.8	2.6	985	19	AAV44439
12	48.8	2.6	985	20	AAZ19349
13	48.8	2.6	985	20	AAZ19137
14	48.4	2.6	12152	22	AAZ08699
15	48	2.6	4403765	22	AAI99683
16	47.8	2.5	24379	18	AAI93095
17	47.8	2.5	24379	19	AAV25925
18	47.6	2.5	30001	18	AAI61016
19	47.6	2.5	30001	20	AAO5110
20	47.6	2.5	114955	20	AAI53491
21	47.2	2.5	1908	16	AAO93364
22	47.2	2.5	1908	20	AAV84065
23	47.2	2.5	1908	21	AAI61403
24	47.2	2.5	77536	21	AAAI4651
25	46.8	2.5	2712	19	AAI94214
26	46.8	2.5	15872	21	AAZ87283
27	46.8	2.5	20394	22	AAI24892
28	46.4	2.5	109519	22	AAI50693
29	46	2.5	44377	18	AAI78508
30	46	2.5	44377	18	AAI80414
31	46	2.5	114955	20	AAI53491
32	45.6	2.4	744	18	AAI74073
33	45.6	2.4	867	22	AAH48828
34	45.4	2.4	31422	21	AAI92302
35	45.4	2.4	31422	22	AAH79278
36	45.2	2.4	1291	20	AAH87940
37	45.2	2.4	2852	19	AAV64558
38	45.2	2.4	2852	19	AAV44449
39	45.2	2.4	2852	20	AAI29359
40	45.2	2.4	2852	20	AAI29147
41	44.8	2.4	2268	21	AAAI0501
42	44.8	2.4	2618	21	AAA38760
43	44.8	2.4	3252	21	AAAI0505
44	44.6	2.4	2376	20	AAI16153
45	44.6	2.4	5224	22	AAI90079

ALIGNMENTS

RESULT 1	AAV72359	Standard: DNA; 1116 BP.
XX	AAV72359;	
AC	XX	
XX	XX	
DT	28-JUL-1999	(first entry)
XX	XX	
DE	Actinomyces sp. 36KD cellulase DNA.	
XX	XX	
KW	Cellulase; detergent; animal feed; nutritional value; textile;	
KW	stone washing; texture modification; appearance; cellulosic fabric;	
KW	pulp; draining; paper; baking additive; starch treatment; grain;	
KW	high-fructose corn syrup production; ethanol production; fibre reduction;	
KW	milling; ss.	
OS	Actinomyces sp.	
XX	XX	
PN	MO9925846-A2.	
XX	XX	
PD	27-MAY-1999.	
XX	XX	
PF	18-NOV-1998;	98WO-US24649.
XX	XX	
PR	24-JUN-1998;	98US-0104308.
PR	19-NOV-1997;	97US-0974041.
PR	19-NOV-1997;	97US-0974042.
XX	XX	
PA	(GENW) GENENCOR INT INC.	
XX	XX	
PI	Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;	
XX	XX	

M. tuberculosis Im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
M. carbonace DNA
Mycobacterium tube
Streptomyces fieno
Streptomyces roseo
Total DNA sequence
S. aureofaciens DN
Human adenosine AI
S. lividans protea
Tripeptidyl aminop
cDNA sequence enco
Nucleotide sequenc
Polyantrum brachys
S. venezuelae vep
Plasmid biosynth
Micromonospora DNA
Platenolide syntha
Human adenosine AI
Myceliophthora the
H. virescens inorg
S. avermitilis ave
Streptomyces averm
Mycobacterium tube
M. tuberculosis Im
M. tuberculosis an
M. tuberculosis re
Non-reducing sacch
Rice 1-deoxy-D-xy
Non-reducing sacch
Mouse Sox1 cDNA.
L05390 cDNA clone.

DR WPI: 1999-347481/29.
P-PSDB; AAY08473.

XX New Actinomycete cellulase useful in detergent compositions, in
PT animal feeds and in treatment of textiles

XX Claim 4; Fig 2; 36pp; English.

XX This invention describes a novel cellulase isolated from an Actinomycete
CC sp. which can be used in detergent compositions, as animal feeds (to
CC increase nutritional value) and in treatment of textiles (e.g. stone
CC washing or modifying texture, feel and/or appearance of cellulosic
CC fabrics, including removal of 'immature' or 'dead' cotton), pulp (to
CC improve draining) and paper. They may also be used as baking additives,
CC for treating starch (in production of high-fructose corn syrup or
CC ethanol) and for treating grain (to reduce fibre during milling).

XX Sequence 1116 BP; 191 A; 432 C; 340 G; 153 T; 0 other:

Query Match 7.0%; Score 131.8; DB 20; Length 1116;

Best Local Similarity 50.7%; Pred. No. 9, 1e-20;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

```

OY 807 ctgagccgagccgacgcgtcgtgctgagcgtggaacgcgcgcatgtgcccggg 866
DB 77 cgtctcccgccagcgaacagatctgcaccgctacgacacacagatccagg 136
OY 867 ggcgtaccgggtgatcaacaacgtatgggcccggagcccgccagtcgatggctg 926
DB 137 accggtactgtgtgcagaacaacgcgtggtggccacgcaccccgctcatcaatgtga 196
OY 927 gactggaaacgggcaacttcacgatcacacgggcgacatcacgaacaacggcgtgg 986
DB 197 ccggcaacggttcgagatcacccacgacggttcggtgcgcgacaaacggcccgca 256
OY 987 ccgcctatccgcatctctcttccgtgctgcacgtgggcgcgcctgcagaaatccggat 1046
DB 257 agtccatctcctcgtctcgaacggttcgacacacgcaacttcgcgcccgacacgc 316
OY 1047 tgcgcgagcgctgcagagcgtgtccgacgtgcgcagcgctggaacgtccaaacgcgatca 1106
DB 317 tgcgcctcgtgatcagctcgatcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 376
OY 1107 cgaacggccgctggaaatgcgcgcacacacacacacacacacacacacacacacacac 1166
DB 377 gcaacggtcgttacaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 430
OY 1167 agcgctaacggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 1226
DB 431 ccaacgggggtgaacccggaacgatcatgtctgttcaacccgggtgcgcgcgcgcgcgcgc 490
OY 1227 cgggcccggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 1286
DB 491 ccatcggtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 550
OY 1287 ccgactggagactggaaatcatcgcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1346
DB 551 gcagcaacggttcgaacgagctgctctcctcctcgtgcgcgcgcgcgcgcgcgcgcgcgcgc 610
OY 1347 agctggaacttgaagcgtccatcgaacgagcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1406
DB 611 gcttcgactlcaagagctcgtctcgaacgagcgttcagcagcgctgcgcgcgcgcgcgcgcgcgc 670
OY 1407 ggtatctgcatcgtgtggaacgagcgttcgaactcgtggagggcgcgcgcgcgcgcgcgcgcg 1466
DB 671 ggtactaccacgcatccacggcggggtcgtcgagcgtggagggcgcgacacgcgtctggccg 730
OY 1467 gcgcgaattttccgtlaacggttga 1491
DB 731 tgaactcgttctcctcgcgcgtga 755

```

RESULT 2
AAZ57029
ID AAZ57029 standard; DNA; 1116 BP.

XX AAZ57029;

XX 19-MAY-2000 (first entry)

XX Actinomycetes cellulase protein encoding DNA.

XX Cellulase; Actinomycetes; detergent; feed additive; textile treatment;

XX pulp; paper; ds.

XX Streptomyces sp.

XX Key

XX CDS

XX sig_peptide

XX mat_peptide

XX WO200009707-A1.

XX 24-FEB-2000.

XX 28-MAY-1999; 99WO-US11971.

XX 24-JUN-1998; 98US-0104308.

XX 18-NOV-1998; 98WO-US24649.

XX 28-MAY-1999; 99US-0321981.

XX (GENEV) GENENCOR INT INC.

XX Jones BF, Van Der Kleij MAH, Van Solingen P, Weyler W;

XX WPI: 2000-224344/19.

XX P-PSDB; AAY67496.

XX A novel Actinomycetes cellulase and related DNA, useful for detergent

XX compositions, treating textiles and paper or pulp

XX Claim 5; Fig 2; 72pp; English.

XX The invention provides a cellulase from Actinomycetes. The cellulase can

XX be used in a detergent composition, as an additive for animal feed and

XX for the treatment of textiles or pulp and paper. The DNA encoding the

XX cellulase can be used to identify homologous cellulases and for

XX recombinant production of cellulases. The present sequence represents a

XX DNA encoding a cellulase from Actinomycetes.

XX Sequence 1116 BP; 191 A; 432 C; 340 G; 153 T; 0 other:

Query Match 7.0%; Score 131.8; DB 21; Length 1116;
Best Local Similarity 50.7%; Pred. No. 9, 1e-20;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

```

OY 807 ctgagccgagccgacgcgtcgtgctgagcgtggaacgcgcgcatgtgcccggg 866
DB 77 cgtctcccgccagcgaacagatctgcaccgctacgacacacagatccagg 136
OY 867 ggcgtaccgggtgatcaacaacgtatgggcccggagcccgccagtcgatggctg 926
DB 137 accggtactgtgtgcagaacaacgcgtggtggccacgcgcaccccgctcatcaatgtga 196
OY 927 gactggaaacgggcaacttcacgatcacacgggcgacatcacgaacaacggcgtgg 986
DB 197 ccggcaacggttcgagatcacccacgacggttcggtgcgcgacaaacggcccgca 256
OY 987 ccgcctatccgcatctctcctcgtgctgcacacgtgggcgcgcctgcagaaatccggat 1046

```



```
OY 1472 gatttttcgtaacggtgca 1491
      || ||| ||||| |
DB 1187 tggttctctctccgcggtgaa 1206

-RESULT 4
AA54513
ID AAA54513 standard; DNA; 1983 BP.
XX
AC AAA54513;
XX
DT 11-APR-2001 (first entry)
XX
DE Fructan exohydrolase (FEH) coding sequence.
XX
KM Fructan exohydrolase; FEH; transgenic plant; recombination;
KM transgene; gene expression; detergent; detergent additive;
KM oral care composition; ds.
XX
OS Zea mays.
XX
PN WO200068402-A1.
XX
PD 16-NOV-2000.
XX
PF 08-MAY-2000; 2000WO-EP04226.
XX
PR 06-MAY-1999; 99BE-0000329.
XX
PA (LEUV-) LEUVEN RES & DEV.
XX
PI Van Den Ende W, Van Laere A, De Roover J, Michiels A;
DR WPI: 2001-007401/01.
XX
PT Novel DNA molecules encoding enzymes having fructan exohydrolase
PT activity for use in transgenic plant production, dental care
PS compositions, and in detergents
XX
PS Claim 9; Fig 4; 45pp; English.
XX
CC Transgenic plants such as Cichorium intybus, Cynara scolymus,
CC Helianthus tuberosus, Scorzonera hispanica, Oryza sativa, Zea mays,
CC Triticum aestivum, Triticum durum, Hordeum vulgare, Secale cereale,
CC Avena sativa, Sorghum vulgare, Phleum pratense, Lolium temulentum,
CC Dactylis glomerata, Pennisetum americanum, Allium cepa, Agave
CC americanum, Agave azul tequilana, Sorghum bicolor and Panicum
CC transformed with a vector encoding a fructan exohydrolase (FEH)
CC enzyme are useful for the recombinant production of FEH or other
CC polypeptides having FEH activity. The FEH polypeptides produced
CC are useful in detergents or as a detergent additive and in oral
CC care compositions.
XX
SQ Sequence 1983 BP; 357 A; 614 C; 683 G; 329 T; 0 other;

Query Match 3.1%; Score 57.6; DB 22; Length 1983;
Best Local Similarity 51.0%; Pred. No. 0.0022;
Matches 213; Conservative 0; Mismatches 199; Indels 6; Gaps 3;

OY 860 gccggggggcgctaccgggtgatcaacaagataggcgcgagagacgccagtgacatt 919
      ||||| ||| || ||| || ||| || ||| || ||| || ||| || ||| || ||| ||
DB 796 gccgggggtgagcaagctgctgaagctcagctgtagtgaacagctccagagcaactacatg 855
OY 920 gaggtcggagctggaaaagggcaactcaagatcaacgagcgccgagatcaagcaaggaac 979
      || || | || | || | || | || | || | || | || | || | || | || | ||
DB 856 gtcggggcggtacgacgacgagcgagcgctgcgtgcgcggagagccgagcggggagac 915
OY 980 aacgtgagccgcatctactcgtgggtgccaatggggcgccgagcgagcaat 1039
      ||||| || || | || | || | || | || | || | || | || | || | || | ||
DB 916 gacgtgagcgagctggcgagcgagctggaactagc---gccacgttgaagcgtccaagaccttc 972
```

```
OY 1040 tcgggattgcgcgcgcgctgcaagagctgtccgac-gtcgcgacagagcttgacgtcac 1098
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 973 ttcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1032
OY 1099 gccgatacgaagagggcgcttgaaatgcgcctcaagacatctgtttagtcccgtaacgaa 1158
      ||||| ||| || | | | | | | | | | | | | | | | | | | | | | | | |
DB 1033 gccgacgacgttcgccaagggctgttcgcgcacgcttcccgcggaagctgtgctgagcgag 1092
OY 1159 ttcgcgcaacggctacagcgcgcgcg--gccgagctgatactgtgcctggaacgcgaacgcg 1216
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1093 gacgcggaagcagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1152
OY 1217 ggcgtgattgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1274
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1153 gtcgtctgcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1210

-RESULT 5
AA14651/C
ID AAA14651 standard; DNA; 77536 BP.
XX
AC AAA14651;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KM FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KM Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KM FK-506 binding protein; polyketide compound; transplant rejection;
KM graft-versus-host disease; uveitis; alopecia universalis;
KM autoimmune chronic active hepatitis; inflammatory bowel disease;
KM multiple sclerosis; primary biliary cirrhosis; scleroderma;
KM neurite outgrowth; nerve regrowth; Parkinson's disease;
KM Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KM peripheral neuropathy; ss.
XX
OS Streptomyces hygroscopicus.
XX
FH Key location/Qualifiers
FT CDS Complement (412..1836)
FT FT /*tag= a
FT FT /*note= "fkbm gene"
FT CDS Complement (2020..3579)
FT FT /*tag= b
FT FT /*note= "fkby gene"
FT CDS 3969..4496
FT FT /*tag= c
FT FT /*note= "fkbr2 gene"
FT CDS Complement (4595..5488)
FT FT /*tag= d
FT FT /*note= "fkbr1 gene"
FT CDS 5601..6818
FT FT /*tag= e
FT FT /*note= "fkbe gene"
FT CDS 6808..8052
FT FT /*tag= f
FT FT /*note= "fkbf gene"
FT CDS 8156..8824
FT FT /*tag= g
FT FT /*note= "fkbg gene"
FT CDS Complement (9122..9883)
FT FT /*tag= h
FT FT /*note= "fkbg gene"
FT CDS Complement (9894..10994)
FT FT /*tag= i
FT FT /*note= "fkbi gene"
FT CDS Complement (10987..11247)
FT FT /*tag= j
FT FT /*note= "fkbj gene"
FT CDS Complement (11244..12092)
FT FT /*tag= k
FT FT /*note= "fkbk gene"
```

FT	CDS	complement (12113..13150)
FT		/tag= l
FT		/note= "fkdb gene"
FT	CDS	complement (13212..23988)
FT		/tag= m
FT		/note= "fkdb gene"
FT	misc_feature	complement (13452..13662)
FT		/tag= n
FT		/note= "ACP6"
FT	misc_feature	complement (13761..14394)
FT		/tag= o
FT		/note= "KR6"
FT	misc_feature	complement (14517..15294)
FT		/tag= p
FT		/note= "ER6"
FT	misc_feature	complement (15438..16587)
FT		/tag= q
FT		/note= "dehydratase domain (DH) 6"
FT	misc_feature	complement (16587..17820)
FT		/tag= r
FT		/note= "acyltransferase domain (AT) 6"
FT	misc_feature	complement (17820..19053)
FT		/tag= s
FT		/note= "KS6"
FT	misc_feature	complement (19116..19326)
FT		/tag= t
FT		/note= "ACP5"
FT	misc_feature	complement (19464..20097)
FT		/tag= u
FT		/note= "KR5"
FT	misc_feature	complement (20241..21420)
FT		/tag= v
FT		/note= "DH5"
FT	misc_feature	complement (21420..22653)
FT		/tag= w
FT		/note= "AT5"
FT	misc_feature	complement (22653..23892)
FT		/tag= x
FT		/note= "KS5"
FT	CDS	complement (23992..46573)
FT		/tag= y
FT		/note= "fkdb gene"
FT	misc_feature	complement (24163..24373)
FT		/tag= z
FT		/note= "ACP4"
FT	misc_feature	complement (24997..26146)
FT		/tag= aa
FT		/note= "DH4 (inactive)"
FT	misc_feature	complement (26146..27430)
FT		/tag= ab
FT		/note= "AT4"
FT	misc_feature	complement (27430..28684)
FT		/tag= ac
FT		/note= "KS4"
FT	misc_feature	complement (28750..28960)
FT		/tag= ad
FT		/note= "ACP3"
FT	misc_feature	complement (29092..29740)
FT		/tag= ae
FT		/note= "KR3"
FT	misc_feature	complement (29869..31018)
FT		/tag= af
FT		/note= "DH3 (inactive)"
FT	misc_feature	complement (31018..32185)
FT		/tag= ag
FT		/note= "AT3"
FT	misc_feature	complement (32185..33439)
FT		/tag= ah
FT		/note= "KS3"
FT	misc_feature	complement (33505..33715)
FT		/tag= ai
FT		/note= "ACP2"
FT	misc_feature	complement (33823..34480)
FT		/tag= aj
FT		/note= "KR2"
FT	misc_feature	complement (34606..35749)
FT		/tag= ak
FT		/note= "DH2 (inactive)"
FT	misc_feature	complement (35749..37144)
FT		/tag= am
FT		/note= "AT2"
FT	misc_feature	complement (37145..38296)
FT		/tag= an
FT		/note= "KS2"
FT	misc_feature	complement (38371..38581)
FT		/tag= ao
FT		/note= "ACP1"
FT	misc_feature	complement (38677..39307)
FT		/tag= ap
FT		/note= "KR1"
FT	misc_feature	complement (39442..40609)
FT		/tag= aq
FT		/note= "DH1"
FT	misc_feature	complement (40609..41842)
FT		/tag= ar
FT		/note= "AT1"
FT	misc_feature	complement (41842..43093)
FT		/tag= as
FT		/note= "KS of extender module 1 (KS1)"
FT	misc_feature	complement (43144..43660)
FT		/tag= at
FT		/note= "ACP of loading domain"
FT	misc_feature	complement (43777..44629)
FT		/tag= au
FT		/note= "ER of loading domain"
FT	misc_feature	complement (44974..46573)
FT		/tag= av
FT		/note= "CoA ligase of loading domain"
FT	CDS	46754..47788
FT		/tag= aw
FT		/note= "fkdb gene"
FT	CDS	47785..52272
FT		/tag= ax
FT		/note= "fkbp gene"
FT	CDS	52275..71465
FT		/tag= ay
FT		/note= "fkba gene"
FT	misc_feature	52362..53576
FT		/tag= az
FT		/note= "KS7"
FT	misc_feature	53577..54716
FT		/tag= ab
FT		/note= "AT7"
FT	misc_feature	54717..55871
FT		/tag= ba
FT		/note= "DH7"
FT	misc_feature	56019..56819
FT		/tag= bc
FT		/note= "ER7"
FT	misc_feature	56943..57575
FT		/tag= bd
FT		/note= "KR7"
FT	misc_feature	57710..57920
FT		/tag= be
FT		/note= "ACP7"
FT	misc_feature	57990..59243
FT		/tag= bf
FT		/note= "KS8"
FT	misc_feature	59244..60398
FT		/tag= bg
FT		/note= "AT8"
FT	misc_feature	60399..61412
FT		/tag= bh
FT		/note= "DH8 (inactive)"

	Query Match	2.9%	Score 53.6:	DB 21:	Length 77536;	
	Best Local Similarity	48.4%:	Pred. No. 0.032:			
	Matches 149;	Conservative 0;	Mismatches 159;	Indels 0;	Gaps 0;	
OY	1156 gaattccgagcaacggctcaccagcgcgcgcccgaactgatgtatgcctgaacttggaacg	1215				
Db	27680 GAAAGTCCAACATCGGCGCACACACTCAGGCCGCCCGGGTGTGCCTCATCAAGATGCT	27621				
OY	1216 cggcgatgatgccgagcgcgagccgcytgcacacgcttgaaacttgccggygccaacttgga	1275				
Db	27620 CCTCCGCATGCGGACAGGCACTTGCCGCGACCTGCACGTAAGAACAGCCGCTCGCA	27561				
OY	1276 agtcgtgatgtcgcgactggtgactgtaataacttgcctcacgsgcgacagacccaacc	1335				
Db	27560 CATTCACTGGAAGAGGCGCGGCGCTGCACCTCTCAACCGATGCCCGCCCTTGCGCGAAGC	27501				
OY	1336 gtccggttgagcagagctcggaaccttaagaaccttacgcagcagcggtccgcggtacaac	1395				
Db	27500 CGACCGCCCCGCGCGCGCGCGGTGTCTCTTCCTTCGCGGTACAGCGGCAACAGCCACGT	27441				
OY	1396 ccgagccgagatgtagtatctgtcatgctgcgttggaagcgggcttcgaaacttggaaggcggagc	1455				
Db	27440 CTGCGTGAAGCCACCGCGCGGAGAGCGCGCGCGGAGAGCGGTGCTCGAAGCC	27381				
OY	1456 cggtctgc 1463					
Db	27380 CGGTGAGC 27373					
RESULT	6					
AAA58471						
ID	AAA58471 standard; DNA; 58857 BP.					
XX	AAA58471;					
XX						
DT	31-OCr-2000 (first entry)					
XX						
DE	Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.					
XX						
KW	BLM gene cluster; bleomycin gene cluster; polyketide metabolite;					
RW	bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;					
KW	thiazoline; bithiazoline; microbial metabolite; sugar; ss.					
XX						
OS	Streptomyces verticillius.					
XX						
FH	Key	Location/Qualifiers				
FT	CDS	223..564				
FT		/tag= a				
FT		/transl_except= (pos: 1..3, aa: Met)				
FT		/note= "ORF 30; encodes AAB07556"				
FT		561..2309				
FT		/tag= b				
FT		/transl_except= (pos: 1..3, aa: Met)				
FT		/note= "ORF 29; encodes AAB07557"				
FT		2767..3486				
FT		/tag= c				
FT		/note= "ORF 28; encodes AAB07558"				
FT		3527..5593				
FT		/tag= d				
FT		/transl_except= (pos: 1..3, aa: Met)				
FT		/note= "ORF 27; encodes AAB07559"				
FT		5806..12294				
FT		/tag= e				
FT		/note= "ORF 26; encodes AAB07560"				
FT		12291..15491				
FT		/tag= f				
FT		/note= "ORF 25; encodes AAB07561"				
FT		15488..21013				
FT		/tag= g				
FT		/note= "ORF 24; encodes AAB07562"				
FT		21010..24666				
FT		/tag= h				

FT		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 23; encodes AAB07563"
ET		24653..32690
ET		/*tag= i
ET	CDS	/note= "ORF 22; encodes AAB07564"
ET		32893..34830
ET		/*tag= j
ET	CDS	/note= "ORF 21; encodes AAB07565"
ET		34827..35804
ET		/*tag= k
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 20; encodes AAB07566"
ET		35818..37302
ET		/*tag= l
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 19; encodes AAB07567"
ET		37299..39215
ET		/*tag= m
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 18; encodes AAB07568"
ET		39301..47181
ET		/*tag= n
ET		/note= "ORF 17; encodes AAB07569"
ET	CDS	47178..49985
ET		/*tag= o
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 16; encodes AAB07570"
ET		49982..51001
ET		/*tag= p
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 15; encodes AAB07571"
ET		50998..52386
ET		/*tag= q
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 14; encodes AAB07572"
ET		52383..52946
ET		/*tag= r
ET		/note= "ORF 13; encodes AAB07573"
ET	CDS	53018..54190
ET		/*tag= s
ET		/note= "ORF 12; encodes AAB07574"
ET	CDS	54187..55824
ET		/*tag= t
ET		/note= "ORF 11; encodes AAB07575"
ET	CDS	55821..56093
ET		/*tag= u
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 10; encodes AAB07576"
ET		56090..57586
ET		/*tag= v
ET		/transl_except= (pos: 1..3, aa: Met)
ET	CDS	/note= "ORF 9; encodes AAB07577"
ET		57583..58857
ET		/*tag= w
ET		/note= "ORF 8; encodes AAB07578"
XX		
PV	WO200040704-A1.	
XX		
PD	13-JUL-2000.	
XX		
PF	06-JAN-2000; 2000WO-US00445.	
XX		
PR	06-JAN-1999; 99US-0115435.	
PR	05-FEB-1999; 99US-0118848.	
PR	05-JAN-2000; 2000US-0477962.	
XX		
PA	(REGC) UNIV CALIFORNIA.	
XX		
PI	Shen B, Du L, Sanchez C, Chen M, Edwards DJ;	
XX		
DR	WPI: 2000-465974/40.	
DR	P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,	
DR	AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,	

ID AAV4439 standard; DNA; 985 BP.
 AC AAV4439;
 XX
 XX
 DT 09-NOV-1998 (first entry)
 DE Mycobacterium tuberculosis antigen XP25 3' DNA.
 XX
 XX
 KM Tuberculosis; infection; diagnosis; antigen; XP25; ss.
 KW
 OS Mycobacterium tuberculosis strain Erdman.
 XX
 PN WO9816645-A2.
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US18214.
 XX
 PR 13-MAR-1997; 97US-0818111.
 PR 11-OCT-1996; 96US-0729622.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-25192/22.
 XX
 PT New isolated Mycobacterium tuberculosis polyepitides and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and diagnosis of tuberculosis
 XX
 PS Claim 4; Page 187-188; 250pp; English.
 XX
 CC This is the 3' region of DNA coding for an antigenic portion of
 CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in
 CC AAV4439. XP25 DNA was isolated from a M. tuberculosis strain Erdman
 CC genomic DNA expression library using sera from patients having
 CC extrapulmonary tuberculosis. It bears no similarity to known
 CC sequences. The invention relates to methods and compositions for
 CC diagnosing tuberculosis. It provides polypeptides (see
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.
 CC tuberculosis antigen, or an immunogenic portion of a M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic
 CC kits for detecting M. tuberculosis infection in a patient using
 CC these polypeptides, antibodies or oligonucleotide probes and
 CC primers.
 CC
 SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other:

Query Match 2.6%; Score 48.8; DB 19; Length 985;
 Best Local Similarity 44.9%; Pred. No. 0.16;
 Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 773 ttcccgatgagcagcaacggaagacccgagcctgagcccgagccgctgagctg 832
 II III IIII
 DB 175 ttccgagtgagcgccgagcagcaagtgagcgccgagccgagccgagcgag 234
 OY 833 tgcgagcgctgagcagcgcgagctgagcgagcgagcgagcgagcgagcgag 852
 III
 DB 235 accgagtaacggcgagcagcgagcgagctgagcgagcagcgagcgagcgag 294
 OY 893 tgggagcgagagagcagcagcgagcgagcgagcgagcgagcgagcgagcgag 952
 IIII
 DB 295 gccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 354
 OY 953 aacagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1012
 II
 DB 355 ggcttgaagcgagcgagcagcagcgagcgagcgagcgagcgagcgagcgagcgag 414

OY 1013 tgcacgtggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1068
 II IIII
 DB 415 atcaacggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 474
 OY 1069 gtccgagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1128
 II
 DB 475 gccgagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 533
 OY 1129 ctacgacatggttgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1188
 I
 DB 534 cggcagcagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 593
 OY 1189 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1248
 II
 DB 594 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 653
 OY 1249 cgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1308
 II
 DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 713
 OY 1309 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 767
 I
 DB 714 cactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag

RESULT 12
 AA219349
 ID AA219349 standard; DNA; 985 BP.
 XX
 AC AA219349;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis antigen 3' XP25 DNA sequence.
 XX
 KM Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KM immunotherapy; diagnosis; immunisation; vaccine; infection;
 KM immune response; skin test; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9942076-A2.
 PD 26-AUG-1999.
 XX
 PF 17-FEB-1999; 99WO-US03268.
 XX
 PR 05-MAY-1998; 98US-0072967.
 PR 18-FEB-1998; 98US-0025197.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527409/44.
 XX
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic
 PT skin tests and protective or therapeutic vaccines or compositions
 XX
 PS Claim 4; Page 171-172; 299pp; English.

The present invention describes polypeptides comprising an immunogenic
 part of a Mycobacterium tuberculosis antigen (Ag). Also described
 are vaccines and fusion protein containing M. tuberculosis Ag's.
 M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 other polypeptides fragments, can be used in pharmaceutical compositions
 or vaccines to generate a protective or therapeutic immune response to
 M. tuberculosis and as reagents in skin tests for diagnosis of
 tuberculosis. Ag can induce proliferation of, or cytokine secretion
 by, T, B or natural killer cells and/or macrophages in
 tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
 AA219225 are used in the exemplification of the present invention.

XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;
SQ

Query Match 2.6%; Score 48.8; DB 20; Length 985;
Best Local Similarity 44.9%; Pred. NO. 0.16; Indels 5; Gaps 2;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```

OY 773 ttcccgatgagcagacgaaggaacccgagcctgagcccgagccagccgctgagctg 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 ttccgagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 234
OY 833 tgcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 accgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294
OY 893 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354
OY 953 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 ggcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 414
OY 1013 tgcacgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474
OY 1069 gtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 533
OY 1129 ctacgacatctggttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
OY 1189 gctgtgacatctggttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
OY 1249 cgtggaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
OY 1309 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 caccatgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 767

```

RESULT 13

AA219137
ID AA219137 standard; DNA; 985 BP.

AC AA219137;

XX 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen DNA encoding 3' Xp25.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;

KW vaccine; immunity; ss.

XX Mycobacterium tuberculosis.

XX MO9942118-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03265.

XX 05-MAY-1998; 98US-0072596.

XX 18-FEB-1998; 98US-0024753.

PA (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;

XX WPT: 1999-527416/44.

DR New polypeptide comprising antigenic portions of M. tuberculosis

PT Claim 4; Page 217; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.

CC Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 2.6%; Score 48.8; DB 20; Length 985;
Best Local Similarity 44.9%; Pred. NO. 0.16; Indels 5; Gaps 2;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```

OY 773 ttcccgatgagcagacgaaggaacccgagcctgagcccgagccagccgctgagctg 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 ttccgagctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 234
OY 833 tgcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 accgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294
OY 893 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354
OY 953 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 ggcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 414
OY 1013 tgcacgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474
OY 1069 gtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 533
OY 1129 ctacgacatctggttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
OY 1189 gctgtgacatctggttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
OY 1249 cgtggaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
OY 1309 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 caccatgacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 767

```

RESULT 14

AA508699/C
ID AA508699 standard; DNA; 12152 BP.

XX AA508699;

XX	15-JAN-2002 (first entry)
PT	
XX	Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
DE	
XX	Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW	variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX	
OS	Mycobacterium tuberculosis.
PN	
XX	US6294328-B1.
XX	
PD	25-SEP-2001.
XX	
PF	24-JUN-1998; 98US-0103840.
XX	
PR	24-JUN-1998; 98US-0103840.
XX	
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Fleischmann RD, White OR, Fraser CM, Venter JC;
DR	WPI; 2001-647261/74.
XX	
PT	Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT	determining the nucleotide sequence of the strain at positions in the
PT	genome corresponding to positions where M. tuberculosis strains CDC
PT	1551 and H37Rv differ -
XX	
PS	Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.
XX	
XX	The invention relates to evaluating strain variation within and between
CC	different populations of the tuberculosis bacterial pathogen,
CC	Mycobacterium tuberculosis or related Mycobacterium by determining the
CC	nucleotide sequence of the first strain at positions in the complete
CC	sequence of the genome that correspond to positions that differ in the
CC	nucleotide sequences of M. tuberculosis strains CDC 1551 (A1919683) and
CC	H37Rv (A1919682). The method is useful for evaluating strain variation of
CC	M. tuberculosis and has valuable application in the fields of
CC	tuberculosis genetics, epidemiology, patient treatment and epidemic
CC	monitoring.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from USPTO
CC	at seqdata.uspto.gov/sequence.html?docid=6294328B1.
XX	
SO	Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

[illegible]

Db 3963482 AACACGTCGGTTCTTCAACTCGGGACAGGCAACCGGGTATCGGCAACTCCAGGAC 3963423
 Oy 1217 ggcgtgatgccgagccgctgagccacgctgtgaactgcccgggccaacttggaa 1276
 Db 3963422 GCGATGTGGGCAATCGGAAATCGGGGGCCACCTGGGGCCCTTGTCGGCGGCAATAC 3963363
 Oy 1277 gctgtatgcccagactgtgactgtattacatgcgc 1312
 Db 3963362 ACCGGTTTCGGGAACCTGGGTCCGTCACACGGGC 3963327

Search completed: August 19, 2002, 13:37:08
Job time: 13327 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:53:16 ; Search time 130.83 Seconds
(without alignments)
3524.070 Million cell updates/sec

Title: US-10-003-759-1

Perfect score: 1877
Sequence: 1 gggcgagtcgaacgctcgggc.....tctccgacacttaaacac 1877

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.8	7.0	1116	4 US-09-104-308-2	Sequence 2, Appl
2	131.8	7.0	1116	4 US-09-321-981-2	Sequence 2, Appl
3	105.8	5.6	1716	4 US-09-321-981-4	Sequence 4, Appl
4	49	2.6	756	1 US-08-642-255-50	Sequence 50, Appl
5	48.8	2.6	985	4 US-09-056-556-182	Sequence 182, Appl
6	48	2.6	4403765	4 US-09-103-840A-2	Sequence 2, Appl
7	47.6	2.5	30001	1 US-08-125-468-1	Sequence 1, Appl
8	47.6	2.5	30001	2 US-08-474-933-1	Sequence 1, Appl
9	47.2	2.5	1908	2 US-08-173-808-1	Sequence 1, Appl
10	47.2	2.5	1908	2 US-08-265-310-1	Sequence 1, Appl
11	47.2	2.5	1908	3 US-08-951-742-1	Sequence 1, Appl
12	46.8	2.5	15872	4 US-09-105-537-1	Sequence 1, Appl
13	46.2	2.5	1288	1 US-08-440-856A-9	Sequence 9, Appl
14	46	2.5	44377	2 US-08-804-227C-7	Sequence 7, Appl
15	46	2.5	44377	2 US-08-804-198-1	Sequence 1, Appl
16	45.2	2.4	1291	3 US-08-997-897-1	Sequence 1, Appl
17	45.2	2.4	1291	4 US-09-156-836B-1	Sequence 1, Appl
18	45.2	2.4	2852	4 US-09-056-556-203	Sequence 203, Appl
19	44.6	2.4	1176	2 US-08-387-942C-17	Sequence 17, Appl
20	44.6	2.4	11230	4 US-09-105-537-32	Sequence 32, Appl
21	44.6	2.4	11230	2 US-08-387-942C-1	Sequence 1, Appl
22	44.6	2.4	36778	4 US-09-105-537-5	Sequence 5, Appl
23	44.6	2.4	38506	3 US-09-320-878-19	Sequence 19, Appl
24	44.4	2.4	2712	3 US-09-025-691-4	Sequence 4, Appl
25	44	2.3	2064	1 US-08-343-428-1	Sequence 1, Appl
26	43.6	2.3	1017	4 US-08-849-751-1	Sequence 1, Appl
27	43.6	2.3	1017	4 US-09-478-816-1	Sequence 1, Appl

28	43.4	2.3	1500	4 US-09-593-711A-10	Sequence 10, Appl
29	43	2.3	1273	4 US-09-319-892-3	Sequence 3, Appl
30	43	2.3	1333	4 US-09-372-422A-9	Sequence 9, Appl
31	42.8	2.3	599	1 US-08-584-226-3	Sequence 3, Appl
32	42.4	2.3	20235	1 US-07-642-734C-3	Sequence 3, Appl
33	42.4	2.3	20235	3 US-08-439-009A-3	Sequence 3, Appl
34	42.2	2.2	1158	4 US-09-372-422A-21	Sequence 21, Appl
35	42	2.2	71989	4 US-09-443-501A-2	Sequence 2, Appl
36	41.8	2.2	432	1 US-08-642-255-48	Sequence 48, Appl
37	41.6	2.2	1155	2 US-08-387-942C-7	Sequence 7, Appl
38	41.4	2.2	669	3 US-08-913-264-2	Sequence 2, Appl
39	41.4	2.2	2563	1 US-08-608-241-5	Sequence 5, Appl
40	41.4	2.2	2563	2 US-08-922-182-5	Sequence 5, Appl
41	41.4	2.2	2563	2 US-08-919-953-5	Sequence 5, Appl
42	41.4	2.2	18994	1 US-08-459-586-4	Sequence 4, Appl
43	41.4	2.2	18994	2 US-08-282-696-4	Sequence 4, Appl
44	41.4	2.2	4411529	4 US-09-103-840A-1	Sequence 1, Appl
45	41	2.2	1182	4 US-09-385-028-19	Sequence 19, Appl

ALIGNMENTS

```
RESULT 1
US-09-104-308-2
; Sequence 2, Application US/09104308
; Patent No. 6187577
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: Van Der Kleij, Wilhelmus A.H.
; APPLICANT: Van Solingen, Piet
; APPLICANT: Weyler, Walter
; TITLE OF INVENTION: No. 6187577el Cellulase Producing Actinomycetes.
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,308
; FILING DATE: 24-JUN-1998
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/974,042
; FILING DATE: 19-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; TELECOMMUNICATION INFORMATION:
; REFERENCE/DOCKET NUMBER: GC539
; TELEPHONE: 650-846-7555
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-104-308-2

Query Match 7.0%; Score 131.8; DB 4; Length 1116;
Best Local Similarity 50.7%; Pred. No. 1.7e-21;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;
```



```

: APPLICANT: Reed, Steven G.
: APPLICANT: Skeiky, Yasir A.M.
: APPLICANT: Dillon, Davin C.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
: NUMBER OF SEQUENCES: 241
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/056,556
: FILING DATE: 07-APR-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: MAKI, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.457
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 182:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 985 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-09-056-556-182

```

Query Match 2.6%; Score 48.8; DB 4; Length 985;
 Best Local Similarity 44.9%; Pred. No. 0.014;
 Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```

Qy 773 ttccccaatggcgacaaagaaacgagccctgagccgagccgagcgtcgaagctg 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 ttgcccgggtggcccccggcgacagtggtggcgccgcccggcgccggcgccg 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 833 tgcgagcgtctggagcgcgcgcgtgtgctggcgggggcgctacccgggtgatcaacaacgt 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 accggtaccggcgacacggcggtgtgtggcgccacggcggtgacagcattcgccggcg 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 893 tggggcgagagacgcgcgcgtgcatgtgagtgctggaatgaacgggcaactcaagatc 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 295 gccggggcgccggcggtgacggcggtgacggcgccacggcggtctcgccgtggtcc 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 953 aacagcgccgacatcacgacaaacaaacgctggcgccctatccgcatctacttcggg 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 355 ggcgtttgacggcgccgacagcgccgacagcgccgacggcgccgacggcgccg 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1013 tgcacatggggcgctgcaagcgcgcgcgtggaatgctggcgccgctgcaagagct 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 atcaacggggcgccggcgccgacgacggcgccgacggcgccgacggcgccg 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1069 gtccgagcgtgagcagagctggaagctcaagcgcgcgcgcgcgcgcgcgcgcgcgcgc 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 gccggcggtctcgcgacaaacggcggtgacggcggtgacggcggtgacggcggtgacggcg 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1129 ctacgacatctgttctacgcttcacgaattccggaacggttaagcgggcgccgga 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 cggcaacggcgccgacggcgccgctcgccctgacacgacacggcgccgacggcgccg 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1189 gctgatgatctggtctgaacttgaaacggcggtgctgctgacggcgccgacggcggtgagc 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 594 gggcaattggcgacacggcgccgctgctgctgacggcgccgacggcgacacgaattttcaa 653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

TREATM

```

Qy 1249 cgtggaactggccggggccacactgggaagtctgtatgcccagctggagctgaattacat 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 cggcgccgacaggtggtggtccggcgccacagcgccgacagcgccggttgccggcgacagc 713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1309 cgccctacggcgacgacgacccacacgctgctggaagcgtgagcttgacctgaagc 1362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 cactgtatgacctgacccgacccgggaaagccgattcacaacagcagcattgcc 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6
 US-09-103-840A-2/c
 Sequence 2, Application US/09103840A
 Patent No. 6294328
 GENERAL INFORMATION:
 APPLICANT: FLEISCHMAN, Robert D.
 APPLICANT: WHITE, Owen R.
 APPLICANT: FRASER, Claire M.
 APPLICANT: VENTER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 CURRENT FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 4403765
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:
 OTHER INFORMATION: CDC 1551
 OTHER INFORMATION: "n" bases at various positions throughout the sequence
 OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Query Match 2.6%; Score 48; DB 4; Length 4403765;
 Best Local Similarity 45.6%; Pred. No. 0.13;
 Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;

```

Qy 860 gccggggggcgccaccggggtgatacaaacgataggggcgagagacgcccagtgatc 919
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963782 gtccgacacgggaacctctcgcaacagaaacttggcagcggcaatracggcaactcgaac 3963723
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 920 gaagtcggaactggaaacgggcaacttcacgatacagcggcgacacgaacggcgaac 979
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963722 gtccgggttcgggaattacggcaacaaactgctggcaccggaacggcgccgacgaac 3963663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 980 aacgtggcgccctatccgcatctacttcgtggtgacgtggcgccgtcgacgaagcaat 1039
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963662 atccggcgccgggaacacggcgacgacgaactgggcttccgacacacggcatttgacac 3963603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1040 tccggaattgcgcgcgcgtgacg--agctgccgaagtgcaagcgttgagcctc 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963602 atccggtttcggaacacacggcgacgacgaactcgttcggcctacccggaacaaacacg 3963543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1097 acgcgacatcaagacggcgccgtggaattgcgcgcgtacgaacatctgttcagctccgc 1156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963542 gtggcgatcggtggtgacacacggcgacgacgaacacacgctgttgaactcggcgacac 3963483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1157 aattccggcaacgctacagcgcgcgcgcgcgagctgatactgctgctgactggaacgcgc 1216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963482 aacaaacgttcggttttcaactccggcaacggcgacacggtatcggaactcagacac 3963423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1217 ggcgtgatcgccggcgccgacgctggaacacggcgacacgctggaactgcccgaactgggaa 1276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963422 gggaaatgtgggaatcggaactcggcgacacacgctggtgctgacggcgccatcattac 3963363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1277 gtctgtatgctgacgtggaactggaattacatcgc 1312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3963362 accgggttcgggaactcggggttcgctcaacacggcg 3963327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT      7
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Pantlin, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08-125,468
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match          2.5%; Score 47.6; DB 1; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0.056;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

OY   904  gacgcgccgagtcattgttagtgcgaactggaaacgggcaacttacgatcacaggcgcca 963
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   25022  GACCGGCAACGGCCTTCGCCCTCGCGCGAGGAGCGCGGAGTGTGCTCGAAGAGCGCGGA 24963

OY   964  tcaccacaacggcaacaacagtgcgcgcataccggccatctacttggygtgcacagtgg 1023
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24963  GCACCGCATCCGGCGCGCGCGGCACATCTACGCCGAGATCACCGGCTAACCCCGGCGCTG 24903

OY   1024  cgcctgcacagacaattcgggaattgcgcgcygcgctgcagaagactgtlccgaactgycgac 1083
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24902  CAACTCTCTACACATGATCAGCGGCTCGCGGCTCGACGGGAAACGAGATGACCAGAGCTATCCG 24843

OY   1084  gagctcgagacgtccacgcgcgatcaagcgcgycgcgctggaatgcccgtctacgaactctggt 1143
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24842  GGTCT---GCGCTCGACCGCAGCGGATGAACCCCGAGAGAGTGTTGGCTACTCTAACGCGCA 24786

OY   1144  cagtcccgtaacgaattccggaacagctacacagcgcgycgcgcgcgaactgtatgtctgct 1203
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24785  CGGCTCGGCGACCAA---GCAGAACGACGCGGCACGAGACGCGCGCTTCAAGAAGAGCCT 24729

OY   1204  gaactcgaaacgcygcgctgtatgctccggygcgycagccgctgtgcccacctggaactgycgcy 1263
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   24728  GGGCGACACCGCTTACCGGGGTGCCGGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 24669

```

```

OY      1264  gaccacactcgggaagcgttgtagtcgacgcacgycgagtctgaattacatcgccaacggygcac 1323
Db      24608  GGGCGGCTGTCCGCCTTGAGTTCGCCGCCACAGCGCCTGGCGCATGAGCACATCGGCTG 24609
OY      1324  gagcgcccaacacgttcgvtgtagcgcgagcttgtaacctgaaaggccttcatcgacgcagcgcgltcg 1383
Db      24608  GCCGCCCAAGCGGCAACCTTCACAGTGGCCGATCCGAGTGCGAGACTGACTGACGTGCCGCT 24549
OY      1384  ccgcggttacatccgagcgcggaagtgatatcgtcatcgcgttgagaagcggcgttcgaactctg 1443
Db      24548  GGTGGCGCGGGAACAGCGGCTGCACAGCTGCTCAACGCTGGCGAGCGCTTCCGCGGCTT 24489
OY      1444  gggaggcgaggcgccgtct 1461
Db      24488  CCAGAGCGGATGTGCT 24471

RESULT      8
US-08-474-933-1/C
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strachy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracyline formation and cosmids
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match          2.5%; Score 47.6; DB 2; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0.056;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

OY      904  gacgcccccagtgcatgagtcgcnagcgaanaacgggcaacttcacgatcacacgggccga 963

```

```

Db 25022 GACCCGCAAGGCTTCGCCCTGGCGAGGGCGCGGATGTCGTGAGAGGACGGGA 24963
Oy 964 tcacgacacaggaacacagctgagccgctatccggcctactctcgtgcaactg999 1023
Db 24962 GCACGGGATCCGGCGGGGCGGACATCTACGGGAGGTACCGGCTACCGCGGGCGTg 24903
Oy 1024 cgcctcagacgaatcgggattcgccgagcgctgcaagagctgctcgaagctg99ac 1083
Db 24902 CAACCTCTACAGATGACGGGGCTGGGTCGACGGGAACGAGATGGCGGACTATCCG 24843
Oy 1084 gacgtgagacccacgacgacacgacgagccgctggaatgcccgtactacacatcgtt 1143
Db 24842 GGTCTGACCGGACCGGATGAACCCGAGGAGTGTGCTACGTCAACCGGA 24786
Oy 1144 cagtcgacgaatcgcgacgacgacgacgacgacgacgacgacgacgacgacgac 1203
Db 24785 CGGCTCGGGACCA---GCAGACGACCGGCGGACGAGACCGCGCTTCAAGAGAGCCT 24729
Oy 1204 gaactggaacgagcgctgacgagcgagcgagcgagcgagcgagcgagcgagcgag 1263
Db 24728 GGGCGACACGCTTACCGGGTGGCGGTGCTGATCAAGTGTGATGAGACACTGCGCT 24669
Oy 1264 ggcacactggaagctgtgtatgacgagcgagcgagcgagcgagcgagcgagcgag 1323
Db 24668 GGGCGGCTGTGCGCTGAGTTCGCCGCAACGCGCTGGCGGAGACACTGCGTGTGT 24609
Oy 1324 gacgacacacgacgctgacgacgacgacgacgacgacgacgacgacgacgacgac 1383
Db 24608 GCGCGGACGCGGACCTGACGAGTGGCGGATCCCGAGTGGAGACTGACGACTGCGCT 24549
Oy 1384 cgcgagctacatcgcgagcgagctgtatctgcatcgtggaagcgagcttgcgaactg 1443
Db 24548 GGTGGGCGGGAACACGCGGTGGGACGAGTGTCAACGCTGGCGGCGCTTCCGCGCTT 24489
Oy 1444 ggaagcgagcgagcgagct 1461
Db 24488 CCAGAGCGGATGCTGCT 24471

```

```

RESULT 9
US-08-173-508-1
; Sequence 1, Application US/08173508
; Patent No. 5616485
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,508
; FILING DATE: 23-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/125/CACO
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1908 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..1759
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 146..148
; OTHER INFORMATION: /note="Met at position -39
; OTHER INFORMATION: represents fmet"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 146..262
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 263..1756
; US-08-173-508-1

Query Match 2.5%; Score 47.2; DB 1; Length 1908;
Best Local Similarity 46.7%; Pred. No. 0.038;
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

Oy 1080 gcaagagctgagacgctacacgacgacgacgacgacgacgacgacgacgacgac 1139
Db 159 GCATACGGGCGGAGGCGGACCGGCTTGGGACGCGCGGACGACACTGTGTACCGCACGCTGA 218
Oy 1140 gttcagctccgtcgaacgaatccggaacgagctacagcgagcgagcgagcgagcgag 1199
Db 219 TCGCCGCGCGCTCTGGGACCGCGGACCGCGGCTTGGGACCGCGGACGACGCGGACGCG 278
Oy 1200 ggcctgacgtgaaacgagcgagctgacgagcgagcgagcgagcgagcgagcgagcgag 1259
Db 279 GCGGAGAGTGGGAGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 338
Oy 1260 cgcgagcgacgtggaagctgtgtatgacgagcgagcgagcgagcgagcgagcgagcgag 1316
Db 339 CGGGCATTCGACTGGGAGGAGCTGGGACCGGACGCGGACGCGGACGCGGACGCGGACGCG 398
Oy 1317 ggcgcagcagcccccacgagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1376
Db 399 GCTACGTCACGGTCCGATGAGTACGCCAAGCCGTACGCGCAAGCATCAGGCTGCGCG 458
Oy 1377 cgttcgcccgcgctacatcgcgagcgagtggtatcgtatcgtatcgtatcgtatcgtatcgt 1436
Db 459 TCGACCGCATTCGCGACACCGGAAACGAGAGCGAGCGGCGGCGGCGGCGGCGGCGGCGG 518
Oy 1437 aactctggaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1472
Db 519 CCGGCGGTCCCGCGGCTCCGCGCTCCGCGCTCCGCGCTCCGCGG 554

RESULT 10
US-08-265-310-1
; Sequence 1, Application US/08265310
; Patent No. 5856166
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; APPLICANT: Malek, Lawrence T.
; APPLICANT: Soosimeyer, Gisela
; APPLICANT: Walczyk, Eva
; APPLICANT: Kryzsmann, Phyllis

```

```

1  APPLICANT: Garven, Sheila
2  TITLE OF INVENTION: STREPTOMYCIN PROTEASES AND IMPROVED
3  TITLE OF INVENTION: STREPTOMYCIN STRAINS FOR EXPRESSION OF PEPTIDES AND
4  TITLE OF INVENTION: POLYPEPTIDES
5  NUMBER OF SEQUENCES: 21
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: Foley & Lardner
8  STREET: 3000 K Street, N.W.
9  CITY: Washington, D.C.
10 COUNTRY: USA
11 ZIP: 20007-5109
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patent Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/265,310
20 FILING DATE: 24-JUN-1994
21 CLASSIFICATION: 435
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/173,508
24 FILING DATE: 23-DEC-1993
25 ATTORNEY/AGENT INFORMATION:
26 NAME: BENT, Stephen A.
27 REGISTRATION NUMBER: 29,768
28 REFERENCE/DOCKET NUMBER: 18740/133/CACO
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 202 672 5300
31 TELEFAX: 202 672 5399
32 TELEX: 904136
33 INFORMATION FOR SEQ ID NO: 1:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 1908 base pairs
36 TYPE: nucleic acid
37 STRANDEDNESS: double
38 TOPOLOGY: linear
39
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 146..1759
43 FEATURE:
44 NAME/KEY: misc_feature
45 LOCATION: 146..148
46 OTHER INFORMATION: /note="Met at position -39
47 OTHER INFORMATION: represents fmet"
48 FEATURE:
49 NAME/KEY: sig_peptide
50 LOCATION: 146..262
51 FEATURE:
52 NAME/KEY: mat_peptide
53 LOCATION: 263..1756
54
55 US-08-265-310-1
56
57 Query Match 2.5%; Score 47.2; DB 2; Length 1908;
58 Best Local Similarity 46.7%; Pred.No. 0.038;
59 Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1.

```

[illegible]

FEATURE:
NAME/KEY: mat-peptide
LOCATION: 263..1756
US-08-951-742-1

Query Match 2.5%; Score 47.2; DB 3; Length 1908;
Best Local Similarity 46.7%; Pred. No. 0.038;
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

QY 1080 gacagagctgagagctcagccgacatacagagcgccgtggaatgcgccttaagacatc 1139
DB 159 GCATACGAGGAGGAGGAGCGCTTGGCCAGCGCCGAGCACTGTGTCACCGCACCTGA 218
QY 1140 ggtcagctccgcagcaaatctccgcaacggtacagcgcgccgagctgatatc 1199
DB 219 TCCGCGGCGCGCTCTCGGCACCGCCGCGAGCCCGCCCGCGGAGCGCCAGCGGCGACG 278
QY 1200 gactgaactgaaagcgcggtgtagtcggcgcgagcgcgctgagccacgctgaaactg 1259
DB 279 GCGGAGAGCTGGAGACCGGAGCGCGCGGTGCCCCCATCCGCCCGCCGCGCGCGCGG 338
QY 1260 ccggggccacctgggaagctgtgtatgcgagctggagctg---aattatagccctacc 1316
DB 339 CGGCGATCGACTGGAGAGCTGCGCAGCGCTGGAACCTGCCCAAGCCCATCGAGTGG 398
QY 1317 ggcgcacagacgcccacacgctcggtgtagcgagctggaagagccctcatcagacgag 1376
DB 399 GCTACGTCACGCTGGCGATGAGTACGCGCAAGCGGTACGCGCAAGCATGAGCTCGCG 458
QY 1377 cggctgcggcggtcacaatccgagcgagtggtatctgcatcgctgagagcgagctcg 1436
DB 459 TCACCGCATCGGACACACCGAAGCAGAGCGAGCGCGCGCTGATCTACAAAC 518
QY 1437 aactctggagggcgggcgctgctggaagcgcg 1472
DB 519 CCGCGGCTCCCGCGGCTCGGCTCGCTTCCTCCCG 554

RESULT 12
US-09-105-537-1
Sequence 1, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match 2.5%; Score 46.8; DB 4; Length 15872;
Best Local Similarity 48.0%; Pred. No. 0.074; Indels 9; Gaps 1;
Matches 172; Conservative 0; Mismatches 177; Indels 9; Gaps 1;

QY 1049 ccggcgcgctgagagctgtcgaagctgagcaagctgagcgtcagccatcacg 1108
DB 7646 cggctgagcggtactgagccggaagctgagcgcgctggaactgacagagctcac 7705
QY 1109 acggcgcgctggaatgcgcgtacagacatcgtgtcagtcacgaatccgagac 1168
DB 7706 aagagactgcggtcgcagcgcttcctctcctcgtctcgcgcatcgcgagac 7765

QY 1169 ggtacagcgcgcgccgagctgatgatctgctgaactggaacgcgcgctgacg 1228
DB 7766 gccgcagagccactacagcgcgccacaacagcgctcgaacgctcgcgcacacgc 7825
QY 1229 ggcgcagcgcgctggtccacagctggaactggcgggccacctgggaagtctgtcc 1288
DB 7826 gccgcagcgcgctggtccgacagctgctgctggcgctctggaagcgagcgacgc 7885
QY 1289 gactggagctggaatcacatcgctac-----ggcgacagcgccaccacgctcg 1339
DB 7886 atggcgagcgctggtcgcgccgagacctcgccgctgagcgcgcggaatcacccg 7945
QY 1340 gtgagcgagctggaactgaagcctcatcagacagcggtcgcccggtacatcc 1397
DB 7946 ctaccccgctgagggcgctcgctcttcgaagcgcgcgctgcgaggaagcgcc 8003

RESULT 13
US-08-440-856A-9
Sequence 9, Application US/08440856A
Patent No. 5750873
GENERAL INFORMATION:
APPLICANT: DELAPORTA, STEPHEN L.
TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVE. N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,856A
FILING DATE: 15-MAY-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: WILDMAN, ROBERT A.
REGISTRATION NUMBER: 36,217
REFERENCE/DOCKET NUMBER: 05463-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1517
TELEFAX: (202) 887-0763
TELEX: 706141
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1288 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-440-856A-9

Query Match 2.5%; Score 46.2; DB 1; Length 1288;
Best Local Similarity 38.5%; Pred. No. 0.059;
Matches 249; Conservative 48; Mismatches 346; Indels 4; Gaps 2;

QY 809 gagccgaagcgaacgctcagctgtgagagctggaacgcgcgatgtgcccggggg 868
DB 339 GTGATCGCGGACATCGACGCGSGCGGAGGCGCTGCGCGCGTGGGSSCGGAS 398
QY 869 cgcctacgggtatcaacaagctatggcgcgaggaacgcccagtgatctgaagtcgga 928
DB 399 GTCRGSTTCGTGGGTCGACAGCTGTGAGGAGAGGACGTGSRGCGCCGTGSGACTGS 458
QY 929 ctggaaacggggaactcaacgatacagggcgcgatacagaacaacggaacgctggcc 988

```

Db 459 GCGCTGTSGCGSRVRCGCGSGS---CTSGACGTSTWSTGCAACACCCGCGGGTGCTG 515
QY 989 gctatccggcatctactctggtgctgacgtggcgctgcaagcaattcgatg 1048
Db 516 GGGCCGCGACCGCGCGCGCGCGCGAGACATCTCTCTTGACAGCAGSSSSAGTTTCAGCCG 575
QY 1049 ccgcgcgcgtgcaagagctgtctcgaactgcaagagctgcaagctcaagcgaatcag 1108
Db 576 GTGCTCCGCTGACGCGCTGCGCGCGCGCGCGCGATGAGACAGCGCGCGCGCGCGATG 635
QY 1109 acgggcgcgtgaaatgctgctcgaacatctgltcagtcctcgatccgaattcgcaac 1168
Db 636 ACCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 694
QY 1169 ggctacaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 1228
Db 695 GCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 754
QY 1229 ggcgcgcgcgcgtgcaagcgctgcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 1288
Db 755 AACGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 814
QY 1289 gactggagctgaattacatgctcctacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1348
Db 815 GTGCGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 874
QY 1349 ctggaacctgaagcgctctacgacgacgacgacgacgacgacgacgacgacgacgacgac 1408
Db 875 GACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
QY 1409 tctctgcatgctgtgagacgagcgctcgaactctggaagcgcgcgcgcgcgcgcgcgcg 1455
Db 935 AAGATGAGAGAGGTGCTCAGGGGCGCTSGCCACGCTCAAGGCGCGCGCGCGCGCGCGCG 981

```

RESULT 14
US-08-804-227C-7/C
Sequence 7, Application US/08804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYMERIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-227C-7

```

Query Match 2.5%; Score 46; DB 2; Length 44377;
Best Local Similarity 47.6%; Pred. No. 0.14;
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```

QY 779 gatgacgacaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaag 838
Db 18968 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18909
QY 839 ccgttggaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 898
Db 18908 GACACGCGCGCTGCTGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18849
QY 899 gcggaagaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 958
Db 18848 GGGGAGTGGGTGCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18789
QY 959 gccgacacgacaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca 1018
Db 18788 GGGGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18729
QY 1019 tggggcgctgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 1064
Db 18728 GGGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18683

```

RESULT 15
US-08-804-198-1/C
Sequence 1, Application US/08804198
Patent No. 5945320
GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rostock, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:

```

1 CLASSIFICATION: 435
2
3 ACTOR/REY/AGENT INFORMATION:
4
5 NAME: CANRELL, PAUL R.
6 REGISTRATION NUMBER: 36,477
7 REFERENCE/DOCKET NUMBER: P
8 TELECOMMUNICATION INFORMATION
9
10 TELEPHONE: 317-276-3885
11
12 INFORMATION FOR SEQ ID NO: 1:
13
14 SEQUENCE CHARACTERISTICS:
15
16 LENGTH: 44377 base pairs
17 TYPE: nucleic acid
18 STRANDEDNESS: single
19
20 TOPOLOGY: linear
21
22 MOLECULE TYPE: DNA (genomic)
23
24 FEATURE:
25
26 NAME/KEY: CDS
27 LOCATION: 350..14002
28
29 FEATURE:
30
31 NAME/KEY: CDS
32 LOCATION: 14046..20036
33
34 FEATURE:
35
36 NAME/KEY: CDS
37 LOCATION: 20110..31284
38
39 FEATURE:
40
41 NAME/KEY: CDS
42 LOCATION: 31329..36071
43
44 FEATURE:
45
46 NAME/KEY: CDS
47 LOCATION: 36155..41830
48
49
50 JS-08-804-198-1

```

Query Match	2.5%	Score 46	DB 2	Length 44377
Best Local Similarity	47.6%	Pred. NO. 0.14		
Matches 136	Conservative 0	Mismatches 150	Indels 0	Gaps 0

[illegible]

```
Search completed: August 19, 2002, 13:28:03
Job time: 12887 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:47:01 ; Search time 3821.57 Seconds

(without alignments)
6629.156 Million cell updates/sec

Title: US-10-003-759-1

Perfect score: 1877

Sequence: 1 gggcgagtcgaacgtcggc.....tctccgaccttaacaac 1877

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estcin:*
4: em_estmu:*
5: em_estlov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_estcl:*
10: gb_estc2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_lnv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.8	3.2	925	12	CNS0091P
2	58.6	3.1	935	12	CNS006XK
3	57.2	3.0	925	12	CNS0091P
4	55.6	3.0	932	12	CNS00720
5	54.6	2.9	932	12	CNS00720
6	54.4	2.9	864	10	BF253875
7	53.8	2.9	964	12	CNS003WG
8	52	2.8	1009	12	CNS010EW
9	51	2.7	844	12	CNS0052P
10	51	2.7	935	12	CNS006XK
11	50.8	2.7	1180	12	AG063851
12	50.6	2.7	700	12	AG060149
13	50.6	2.7	878	10	BE785566
14	50.4	2.7	1022	9	BE040131
15	50	2.7	894	12	AG060214
16	49.8	2.7	449	10	BE517742
17	49.6	2.6	682	9	AU095988

C 18	49.2	2.6	534	12	A2934730
C 19	49.2	2.6	546	12	A2935386
C 20	49.2	2.6	644	9	AU030637
C 21	48.2	2.6	443	9	AV914669
C 22	48.2	2.6	446	10	BM373545
C 23	48	2.6	559	10	BE405069
C 24	48	2.6	588	10	BE253157
C 25	48	2.6	604	9	AV941745
C 26	48	2.6	606	10	BF255408
C 27	48	2.6	631	9	AL505254
C 28	48	2.6	635	9	AV932837
C 29	48	2.6	637	9	AV933847
C 30	48	2.6	666	10	BF628036
C 31	48	2.6	719	9	AV935472
C 32	48	2.6	830	10	BF254391
C 33	48	2.6	1101	12	CNS0175Y
C 34	47.8	2.5	307	9	AU070991
C 35	47.8	2.5	578	9	AU094712
C 36	47.8	2.5	623	9	AU032635
C 37	47.6	2.5	517	9	AV933848
C 38	47.4	2.5	407	10	BE475922
C 39	47.2	2.5	637	10	BE484578
C 40	47	2.5	657	10	BI717319
C 41	47	2.5	677	9	AL505169
C 42	46.8	2.5	269	10	BE636900
C 43	46.8	2.5	641	9	AU093820
C 44	46.8	2.5	645	12	CNS01213
C 45	46.8	2.5	669	9	AU096005

ALIGNMENTS

RESULT 1
LOCUS CNS0091P
DEFINITION Drosophila melanogaster genome survey sequence Tregi end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

BACRI4B09 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.
 AL066742
 VERSION AL066742.1 GI:4945205
 KEYWORDS GSI.
 SOURCE Fruit fly.
 ORGANISM *Drosophila melanogaster*
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; *Drosophila*.
 1 (bases 1 to 932)
 Genoscope.
 REFERENCE Direct Submission
 AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 JOURNAL - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley *Drosophila* Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 Location/Qualifiers
 source 1..932
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone.lib="RPCI-98"
 /clone="BACRI4B09"
 /note="end : 17"
 BASE COUNT 155 a 202 c 241 g 91 t 243 others
 ORIGIN
 Query Match 2.9%; Score 54.6; DB 12; Length 932;
 Best Local Similarity 30.8%; Pred. No. 1.8; Indels 0; Gaps 0;
 Matches 116; Conservative 81; Mismatches 180; Indels 0;

RESULT 6
 LOCUS BF253875
 DEFINITION
 BF253875 864 bp mRNA linear EST 22-OCT-2001
 HVSMEF0002F20f Hordeum vulgare seedling root EST library HVCDA0007
 (Etioolated and unstressed) Hordeum vulgare cDNA clone
 HVSMEF0002F20f, mRNA sequence.
 ACCESSION BF253875
 VERSION BF253875.2 GI:11116780
 KEYWORDS EST.
 SOURCE Barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae
 1 (bases 1 to 864)
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library unpublished (2001)
 On Nov. 16, 2000 this sequence version replaced gi:11182980.
 Contact: Wing, RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 333
 Seq primer: AATTAACCTCACTAAAGG
 High quality sequence stop: 776.
 Location/Qualifiers
 source 1..864
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEF002F20f"
 /clone.lib="Hordeum vulgare seedling root EST library
 HVCDA0007 (Etioolated and unstressed)"
 /tissue.type="Seedling root"
 /lab_host="TJCI21"
 /note="Vector: LambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give plasmid SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"
 BASE COUNT 172 a 260 c 271 g 161 t
 ORIGIN

Query Match	2.9%	Score 54.4	DB 10	Length 864
Best Local Similarity	48.1%	Pred. No. 1.9		
Matches 154	Conservative 0	Mismatches 166	Indels 0	Gaps 0
0y	1091	acgccacgcgcgcatcacagcggccctcgtgaatgcgcctcaacacatcgtgtcagtc	1150	
Db	195	ACCGGATGCGAAGACCCGCGCGAGCTCGACGACGCGCTTGTGAATGTGACACAGGCTG	254	
0y	1151	gtcacgacttcgcgcaacgcgctacagcgcgcgcgcgaagtctgatctcgtgtaactgg	1210	
Db	255	AGGAAGAAGTGTGCTGCGCGAGAGCTCGCAACAGACGAGCGCTCAAGCTCTGGGAATCAGG	314	
0y	1211	aacggcgcgcgtgatcgcggcgcgcgccgcgcgtgtgcacccgctggaacttgcggcgccacc	1270	
Db	315	GACCTGGTACGACGGACCTTGACCCCACTGGCCGACGCTGGCTGCAAGCGCGGAAGAGC	374	
0y	1271	tgggaagctcgtgatcgcgcgcgcgcgcgcgaatcacatgcctacgcgcgcgaacgagcc	1330	
Db	375	TTCCAGTTCTGGACCCACGACGATGGGAGAACACGCGACCTGCTCCAACTGGACCAAGCAC	434	
0y	1331	accacgctcgtgtgagcgaagcttggaacctgaagcgtcttcacatcgacgaacgcgcgcgcgc	1390	
Db	435	GCGTACTTCGCGACGCGCGCTGCTTCAAGCGCCGCCCAACACTTCATCCTCGCGC	494	
0y	1391	tacatccgcgcgcgagtgata 1410		
Db	495	GACGCGGGATCGTCCGTA 514		
RESULT 7				
CNS003MG/c				
LOCUS				
DEFINITION	CNS003MG	964 bp	DNA	linear GSS 03-JUN-1999
DESCRIPTION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR09E09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL065254			
VERSION	AL065254.1	GI:4942606		
KEYWORDS	GSS.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
AUTHORS	1 (bases 1 to 964)			
TITLE	Genoscope.			
JOURNAL	Direct Submission			
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)			
	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).			
	The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .			
FEATURES	Location/Qualifiers			
SOURCE	1..964			
	/organism="Drosophila melanogaster"			
	/db_xref="taxon:7227"			
	/clone_lib="RPCI-98"			
	/clone="BACR09E09"			
	/note="end : TET3"			
BASE COUNT	234 a	160 c	60 g	105 t 405 others
ORIGIN				

```

Query Match          2.9% ; Score 53.8; DB 12; Length 964;
Best Local Similarity 21.4% ; Pred. No. 2.5;
Matches   91; Conservative 154; Mismatches 180; Indels    1; Gaps    1;

QY      1046 ttgcgagcgacgctgtcagaagtcttcgaactgtgcacaagatcttgaaacctcaaccggcatc 1105
        |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       954 TTGCBSGCGCYKTTTCSBSCGGSGSTBYSCSSTSACCBXGSBGSGSBDBTGSSGST 895

QY      1106 acgacaggccgccctggaatggccgcccctaagacatttgttccagtcccgcttaagaattccggc 1165
        :::::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db       894 SGTSTTYHBCSVSGSSSTRGSSGBSRCCSCCTCBBSSCSGCCGCCCKCBMTTTSMS 835

QY      1166aacgactcacagcgcgagccgcagactgtgatctgtcctgaactggaagcgagcgctgatat 1225
        |:::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db       834 ANMGSSSSGSGSGSSGSGSVSRTRKKRRSSRSGSGSGKGSGGRAGAGCGGGCGGCRNG 775

QY      1226 ccggagcgagcagccgcgtgtgcacacgtltgaactgyccggagccaacctyggaaagtlctgttat 1285
        ::::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db       774 SSKSTSGTCGSGSGGKRRTTTRTKGSSSSGCGTTGSCSGGGAASGSSGRCVGABRBKTGT 715

QY      1286 gcccagcttyggagctgaataatcatactgcctccaccggcgcaagacccccaccaagtcgcttagac 1345
        Db       714 GSTTGSKVKSXTTTTSGKVNTSSRSXYRSTRYSGARCTTBDSAAAAGNS--SAAMSMKC 656

QY      1346 gagctygacactgaaaggcccttacatcgaacagcgcgtgcgccgcggtcatataccgcgcgagay 1405
        Db       655 SASVSIVGVSCCMCASASSCCSSSCMVCGCGKCGFKCCSSSVSRSGSPTGCGVCGSGSG 596

QY      1406 tttgatctctgcatgcggtgtgagaaacgggtcttcgaactctytggaggcggggacggtctcgga 1465
        Db       595 SGBTSTSSGGSVAANAACAACACGCAKCGAGSSVSGAACCGYTWSSRADCAAAGAAGRVRVSSGTVA 536

QY      1466 agagcc 1471
        Db       535 VRBMTH 530


RESULT      8 CNS101EM/c     1009 bp DNA linear GSS 26-JUL-1999
LOCUS       Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BAON03P19 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
VERSION     ALTO98882
KEYWORDS    ATO98882.1 GI:5610493
SOURCE      GSS.
ORGANISM   fruit fly.
            Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1009)
AUTHORS    Genoscope.
TITLE       Direct Submission
JOURNAL    Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr

COMMENT     - Web : www.genoscope.cns.fr
             Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP)-http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES         Location/Qualifiers
source           1..1009
                 /organism="Drosophila melanogaster"
                 /plasmid="pBelOBAC11"
                 /db_xref="taxon:7227"
                 /clone_lib="DrosBAC"
```


COMMENT	BASE COUNT	ORIGIN
<p>Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammaster in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila-bac.htm.</p>	257 a	170 c 162 g 96 t 250 others
<p>FEATURES</p> <p>source</p> <p>Location/Qualifiers</p> <p>1..933</p> <p>/organism="Drosophila melanogaster"</p> <p>/db_xref="taxon:7227"</p> <p>/clone_lib="RPCI-98"</p> <p>/clone="BACRI4N09"</p> <p>/note="end : 17"</p>		

Query Match	2.7%:	Score 51:	DB 12:	Length 935:
Best Local Similarity	28.4%:	Pred. No. 8.2:		
Matches 120:	Conservative 95:	Mismatches 208:	Indels 0:	Gaps 0:
QY 1286	gcccactlgygaactlgygaattacatcgctccactcggcgacagcagccaccacagtlcgltgagc	1345		
Db 510	GSCSMCRFSSGSGTGCSTCCMYMSSSVSCCSCGRCGXCSTSCMKMCSCTYGCXGCGC	569		
QY 1346	gagctlgyaacctcgaagcgcccttcatalcgacagcagcgggtcgcgcgcgcgtcatatccgcgcgag	1405		
Db 570	GCSFSCSSSSCCSMBSTSTFCSTBCTKCCSSGCSSTGSCGTGCCGGGSGCGCGSGCGCGC	629		
QY 1406	lgtatctctgcacatgcgtlgyagacagcgcttcgaactctctggaagcggcgccgcgtctgcga	1465		
Db 630	GSGGCGGCGGCGSSSGSGSSGSSGSSGSSGSGSGSGSGSGSGSGCGCGCGCGCGSGSG	689		
QY 1466	agcgccgaatttttcgtataagcgtlgtcgtatgcctctgtcacaagcggcaccagcgtatagccaga	1525		
Db 690	CGSCGCGSSGCGSGCGCGSGCGSGCGSCGSCGCGGCGSGSGSGCGSCCGCGCGCGCGSS	749		
QY 1526	gaagcaccccgctcgggcgacctatgcggcgcgccgtcgcgtcttgcgtcacttgcgttgcgtgc	1585		
Db 750	SSGSGCSCGCGCGCMSCGCGCGSSGCGCGSCGCGCGSCCGCGCGCGCGCGCGCGSGC	809		
QY 1586	gctacagagacgctlcaagccgagcgtlgtcgtlctcccttgaagcgcctlgtlctgcgcga	1645		
Db 810	GGCGCGGSSGSGSGSCCGSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGSGG	869		
QY 1646	cgagaactctatccgaaaagcgaagcagagacgtgttaagccagggccttgcactgcgcac	1705		
Db 870	CGSGSSGGGSSSSGSGSGCGCGSGSGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	929		
QY 1706	cag 1708			
Db 930	CCG 932			
RESULT 11				
LOCUS	AG063851/c	1180 bp	DNA	linear
DEFINITION	Pan troglodytes DNA, clone: PTB-052M02.F, genomic survey sequence.			
ACCESSION	AG063851			
VERSION	AG063851.1	GI:16615653		
KEYWORDS	GSS; GSS (genome survey sequence).			
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male			
ORGANISM	Pan troglodytes			

REFERENCE AUTHORS	TITLE JOURNAL REFERENCE AUTHORS	TITLE JOURNAL
Eukaryota: Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.	1 (sites) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., rotoki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of library PTB Unpublished 2 (bases 1 to 1180) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., rotoki, Y., Watanabe, H. and Sakaki, Y.
	Direct Submission Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenhiro chou, Tsurumi ku, Yokohama, Kanagawa 220-0045, Japan (E-mail: chunpbes@sc.riken.go.jp, URL: http://nsp.gsc.riken.go.jp/, Tel: 81-43-503-9111, Fax: 81-43-503-9170)	
COMMENT	Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.	
PRIMERS		

FEATURES		Location/Qualifiers
Source	1. .1180	/organism="Pan troglodytes" /db_xref="taxon:9598" /clone="PRB-052M02.F" /sex="male" /cell_type="lymphoblast" /clone_idb="PRB Chimpanzee Male BAC Library"
BASE COUNT	182 a 511 c 453 g 28 t	6 others
ORIGIN		
Query Match	2.7%; Score 50.8; DB 12; Length 1180;	
Best Local Similarity	46.5%; Pred. No. 9.2;	
Matches 198; Conservative	0; Mismatches 227; Indels 1; Gaps 1;	
OY	171 ggcgcacatgtgcgcgctcgtcctcatcagtggtatgtgtcgtgacacacagatgtggtgcgagac	230
Db	712 GCGCCCGCGCGGCACATCTGCTCAAGCGGGGCGCCCGCGCTACGGCGGTGGCGGCGGT	653
OY	231 cacacaactacatccctgcgcgcacaaacgtgaagttgctgcgagcgtctgcctcctcgtgcgtcgtg	290
Db	652 GCGCGCGGT-GGTGGCGGCGCGGCGTGGGTGTGTGGCGCGGCGGTGCTGTGACAGTCCGGC	594
OY	291 ccaacgactacagcgtctccggagttgtgtctactcctccgtcccatcctagaacgcatgtgaac	350
Db	593 GTGGCGGCTCCGCTCCCGCCGCGGCGGCGGTGCTGCGCCCGCCCTTCCTCTCCAG	534
OY	351 gtgagatgtgcgcgcctgtgcgtctatcttcctgcctatgcgcgaaggggagactgtgctctgcgc	410
Db	533 CGCGGGGCGGCGCGCGCTCCGCGGCTGCTGCTCTGCGCGGTGGGTGCTGCTGCTGCGC	474
OY	411 cgcgcgcagcgcgcgcgcgagttcgtgaltcgtgcggagacgcggaaccgtgtcttccaccgcgcgc	470
Db	473 GCGCTGCTGGGCGTGGGCGGCGCGCGGCTGGCGGTGGCGGGGGGTCTCCGCGTCCGGGCGC	414
OY	471 gggtagcaggtgcgaagcgcaaacgcgtgtgtgtgtgagtcgcgcgcgtctccgatccgaacg	530
Db	413 TTGTGCGGCGCGGCGGCGCGCGCGCGCGCTTGTCTTCTGCGTGTGGCGGCGCGTGGGCGC	354
OY	531 cgcgtgcctatgtgtcgtgttccaatacgcgcgcacacgcacgctgttcaaccgggcgcgcgtcgc	590
Db	353 CGCGGGGCGCGCGCTGCGCGCGCGCTCCCGCGCGGCGGCGGCGGCGCGCGCGCGCTCG	294
OY	591 cgcgcctt 596	
Db	293 CCCGTT 288	


```

RESULT 12
AG060149          700 bp  DNA      linear  GSS 03-NOV-2001
LOCUS             Pan troglodytes DNA, clone: PTB-047122.R, genomic survey sequence.
DEFINITION        AG060149
ACCESSION         AG060149
VERSION           AG060149.1 GI:16611379
KEYWORDS          GSS: GSS (genome survey sequence).
SOURCE            Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
                  BAC library clone:PTB-047122.R.
ORGANISM          Pan troglodytes
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;
REFERENCE         Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
                  1 (sites)
AUTHORS           Tokoki,Y., Watanabe,H. and Sakaki,Y.
TITLE             BAC end sequences of library PTB
JOURNAL           Unpublished
REFERENCE         2 (bases 1 to 700)
AUTHORS           Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
                  Tokoki,Y., Watanabe,H. and Sakaki,Y.
TITLE             Direct Submission
JOURNAL           Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
                  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
                  1-7-22 Suehiro-cho,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
                  (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
                  Tel:81-45-503-9111, Fax:81-45-503-9170)
                  Clones are derived from the chimpanzee BAC library PTB. This BAC end
                  was generated during the R&D process and may have higher chance of
                  clone tracking errors.
COMMENT           PRIMERS
                  Sequencing: M13Rev
LIBRARY           Vector : pKS145
                  R.Site 1 : SacI
                  R.Site 2 : SacI
FEATURES          Location/Qualifiers
                  1..700
                   /organism="Pan troglodytes"
                   /db_xref="taxon:9598"
                   /clone="PTB-047122.R"
                   /sex="male"
                   /cell_type="lymphoblast"
                   /clone_11b="PTB Chimpanzee Male BAC library"
BASE COUNT       30 a 291 c 334 g 15 t 30 others
ORIGIN
Query Match      2.7%; Score 50.6; DB 12; Length 700;
Best Local Similarity 42.7%; Pred. No. 9.4;
Matches 243; Conservative 0; Mismatches 323; Indels 3; Gaps 1;
QY 776 ccccatgacgaacgaagaaacccgagccctgagcccgacccgtcgactgtgc 835
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 24 cccgngtgnnnnggannngagngcccgagggggggggcgccggttgggg 83
QY 836 ggaagcttgagcgcgcgcatgtggccgcgggggcgtacacgggtgataacaacgtc 895
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 84 GGGNNGGNGGGGGAGATTGGGGCGNGGGGGGAGAGGGGGGGGGGGGGGGGGGG 143
QY 896 gggcgagagacgcgcagtgatgtgagtgctgactggaacgggcgaacttcagatca 955
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 144 GCCGGGGGNGGGGCGCGGGGGCGCGGGGGGTTTCGGGGGGGGCGCGGNNAGG 203
QY 956 cgggcgcgtacgaacaaacggaacagctggcgcgtatccgycacatctacttcggtgc 1015
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 204 GGGGCGGAGACAGACACACACTCGTGGGTATCAAGGGCGGGGCCCCGCGGNGCC 263
QY 1016 cactgagggccttgcacaggaacatctggatgtcgcgcggtgtcaggaagctgtccgac 1075
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 264 CGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
QY 1076 gtgcgacagagctggaagctcaagcgcgatcaagagggcgcgctggaatgcgcgtaagac 1135

```

```

RESULT 13
BG785566/c       878 bp  mRNA      linear  EST 20-MAY-2001
LOCUS             SEA00C005523 Sea urchin primary mesenchyme cell cDNA library,
DEFINITION        Strongylocentrotus purpuratus cDNA clone PC_0026_A1.G09_MR 5', mRNA
                  sequence.
ACCESSION         BG785566
VERSION           BG785566.1 GI:14156579
KEYWORDS          EST.
SOURCE            Strongylocentrotus purpuratus.
ORGANISM          Strongylocentrotus purpuratus
                  Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
                  Echinioidea; Euechinoidea; Echinoidea;
                  Strongylocentrotidae; Strongylocentrotus.
REFERENCE         Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
                  Ettensohn,C.A.
                  A large scale analysis of mRNAs expressed by primary mesenchyme
                  cells of the sea urchin embryo
JOURNAL           Development 128 (13), 2615-2627 (2001)
MEDLINE           21384984
COMMENT           Contact: Ettensohn CA
                  Dept. Biol. Sci.
                  Carnegie Mellon University
                  4400 Fifth Avenue, Pittsburgh, PA 15213, USA
                  Tel: +1 412 268 5849
                  Email: ettensohn@drew.cmu.edu.
FEATURES          Location/Qualifiers
                  1..878
                   /organism="Strongylocentrotus purpuratus"
                   /db_xref="taxon:7668"
                   /clone="PC_0026_A1.G09_MR"
                   /clone_11b="Sea urchin primary mesenchyme cell cDNA
                   library"
                   /tissue_type="embryo"
                   /cell_type="primary mesenchyme cells"
                   /lab_host="E.coli"
                   /note="Vector: pSPORT1; Site.1: NotI; Site.2: SalI; oligo
                   dt priming from poly A+ RNA, directionally cloned"
BASE COUNT       12 a 396 c 407 g 57 t 6 others
ORIGIN
Query Match      2.7%; Score 50.6; DB 10; Length 878;
Best Local Similarity 46.7%; Pred. No. 9.7;
Matches 188; Conservative 0; Mismatches 213; Indels 2; Gaps 1;
QY 244 ccgagcgacaaacgtgaagtgagcagcgccctgtggcgtgtggcacaagcgtc 303
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 711 CCGCCCGCGCGCGCGGAAGAGCACCGCGCGGGGGGGCGCGCGCGCGCGCGCG 652
QY 304 ctccgcgatgtgtctactcgcgtlccatctaaagacatgaaagtgagtgccg 363
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |

```


[illegible]

RESULT 14	
BE040131/c	
LOCUS	BE040131
DEFINITION	1022 bp mRNA linear EST 07-JUN-2000
DESCRIPTION	OD102C08 OD <i>Oryza sativa</i> cDNA 5' similar to gata-binding transcription factors, mRNA sequence.

ACCESSION	BE040131
VERSION	BE040131.1
	GI:8335147

KEYWORDS	EST.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa

REFERENCE
1 (bases 1 to 1022)
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretaceae; Oryzaeae; Oryza.
1 (bases 1 to 1022)

AUTHORS Bonnett, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferrer, H., Kawasaki, S., McCollough, A., Michalowski, C.B., Palacio, C., Scera, G., Wheeler, M., and Zepeda, G.R.
TITLE Functional Genomics of Plant Stress Tolerance
JOURNAL Unpublished (2000)
COMMENT Contact: Michalowski, C.B.

FEATURES
source
University of Arizona
Bio Sciences West Room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cdm@u.arizona.edu
An open reading frame exists.
Location/Qualifiers
1..1022

BASE COUNT	ORIGIN
196 a	285 c
292 g	217 t
32 others	

Query Match	2.7%	Score 50.4	DB 9	Length 1022
Best Local Similarity	55.0%	Pred. No. 11		
Matches 99	Conservative	0	Mismatches 81	Indels 0
				Gaps 0

[illegible]

D

Qy Db	<div style="position: relative; height: 100px;"><div style="position: absolute; left: -10px; top: 0; bottom: 0; width: 1px; background-color: black;"></div><div style="position: absolute; right: -10px; top: 0; bottom: 0; width: 1px; background-color: black;"></div><div style="position: absolute; left: 50%; top: 0; bottom: 0; width: 1px; background-color: black; transform: translateX(-50%);"></div></div>	1271 tgggaagtcgtgtatgcacgacttggacttggaattcatalcgcttaaccgagcaagaagccc 1330 363 ttccggcggcggcgtctggttgggtggcgcaggagacgtccattctcttTTGGCGCCGCAACTTC 304
----------	--	---

```
QY      1331 accacgctcggtgatgcagcagccggaccctgaaggacctatcgaaagcaagcygtcgccccgcgcg   1390
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       303 ACCACCTTCGGTGGGAGTTGCAGTGCTTGGCACCAACAAGAAGCTCGGGCTCCGCCGCCGC    244
```

RESULT	15
LOCUS	AG060214
DEFINITION	AG060214 Pan troglodytes DNA, clone: PTB-047K07.F, genomic survey sequence.
ACCESSION	AG060214
VERSION	AG060214.1 GI:16611446
KEYWORDS	GSS: GSS (genome survey sequence).
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
ORGANISM	BAC library clone: PTB-047K07.F.
	Pan troglodytes

REFERENCE
1 (sites)
Eukaryota: Metazoa: Choriata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Hominae: Pan.

AUTHORS	TITLE	REMARKS
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of Library PTB	

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 894)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: cfrimbes@gscc.riken.go.jp, URL: <http://npg.gscc.riken.go.jp/>, Tel.:81-45-503-9111, Fax:81-45-503-9170). Clones are derived from the chimpanzee BAC library PNB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

```

Sequencing: -21M13
LIBRARY
Vector      : pRS145
R.Site 1    : SacI
R.Site 2    : SacI.
Location/Qualifiers
1. .894
FEATURES
source

```

```

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-047K07.F"
/sex="male"
/cell_type="lymphoblast"
/clone_id="PTB Chimpanzee Male BAC Library"
BASE COUNT      83 a      380 c      398 g      7 t
ORIGIN

```

Query Match	2.7%	Score 50;	DB 12;	Length 894;
Best Local Similarity	44.0%	Pred. NO. 12;		
Matches 307; Conservative	0;	Mismatches 383;	Indels 8;	Gaps 3;

Oy ggcgcaacacggaagaaaggaacccggagcctgagaccgcagaccgcgtctgatgcyagaagc 84
 | | | | | | | | | | | | | | | | | | |
Dd 1-3 gggggcannaggg 20

Oy	842	tggagcgcgcccgaatgtgcccggggcgctaccgggtgatcaacaacgtattgggcgcg	901
Db	203	cggggggggggcgagacacgacggcgacacagggagcAAAAAGGACAAACACTGGGTGGC	262

Dy 902 gagaacgcgcagtcattgagtgctgacactgtaacgcygcaattcacatatacagggcc 961
 263 gattccaccncggcccgcgccccggccccggccgccccggcccaaggcggaacgggac 322

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:46:46 ; Search time 53.28 Seconds
(without alignments)
544.112 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446
Sequence: 1 MNVNRALVLSLLFLFGCDW.....FLMRGAGLRADSPSVTVQ 261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /A_Geneseq_032802.*
2: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.*
3: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.*
4: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.*
5: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.*
6: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.*
7: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.*
8: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.*
9: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.*
10: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1318.5	91.2	260	AAV06369	Rhodothermus marin
2	1318.5	91.2	260	AA14882	Emericella deserto
3	1318.5	91.2	260	AA14882	Amino acid sequenc
4	470.5	32.5	105	AAV06342	Rhodothermus marin
5	400.5	27.7	386	AAV67497	Protein sequence o
6	386.5	27.4	371	AAV06367	Streptomyces sp. E
7	386.5	27.4	371	AAV08473	Actinomyces sp. 3
8	396.5	27.4	371	AA14880	Streptomyces livid
9	396.5	27.4	371	AAV64345	Amino acid sequenc
10	396.5	27.4	371	AAV67496	Actinomyces cell
11	343	23.7	381	AAV06368	Streptomyces livid

12	343	23.7	381	21	AA14881	Rhodothermus marin
13	330.5	22.9	429	21	AAV84346	Amino acid sequenc
14	255.5	17.7	261	20	AAV88462	Bacillus lichenifo
15	252.5	17.5	264	20	AAV06370	Erwinia carotovora
16	252.5	17.5	264	21	AA14883	Streptomyces livid
17	252.5	17.5	264	21	AAV84348	Amino acid sequenc
18	224.5	15.5	259	20	AAV06351	Aspergillus aculea
19	224.5	15.5	259	21	AA14864	Aspergillus aculea
20	224.5	15.5	259	21	AAV84329	Amino acid sequenc
21	214	14.8	239	21	AAV06353	Aspergillus kawach
22	214	14.8	239	21	AAV84331	Aspergillus kawach
23	214	14.8	239	21	AAV84331	Amino acid sequenc
24	213.5	14.8	254	21	AA14869	Humicola insolens
25	213.5	14.8	254	21	AAV84334	Amino acid sequenc
26	210.5	14.6	246	20	AAV06356	Emericella deserto
27	210.5	14.6	246	21	AA14879	Actinomyces 11A68
28	210.5	14.6	246	21	AAV84344	Amino acid sequenc
29	210.5	14.6	254	21	AA14868	Humicola grisea EG
30	210.5	14.6	254	21	AAV84333	Amino acid sequenc
31	210.5	14.6	254	22	AAU07558	Humicola grisea en
32	210.5	14.6	254	22	AAU07584	Humicola grisea en
33	210	14.5	254	22	AAU07583	Humicola grisea en
34	203	14.0	255	20	AAV06356	Humicola insolens
35	200.5	13.9	348	20	AAV06362	Gliocladium roseum
36	200.5	13.9	348	21	AA14875	Gliocladium roseum
37	200.5	13.9	348	21	AAV84340	Amino acid sequenc
38	199	13.4	104	20	AAV06357	Streptomyces sp. E
39	194	13.4	253	20	AAV06355	Humicola grisea EG
40	191.5	13.2	234	22	AAU07582	Trichoderma reesei
41	186.5	12.9	244	19	AAV86593	Tiarsoporella phas
42	186.5	12.9	244	21	AAV44341	T. phaseolina xyl
43	186	12.9	244	20	AAV06359	Fusarium javanicum
44	186	12.9	244	21	AA14872	Amino acid sequenc
45	186	12.9	244	21	AAV84337	Amino acid sequenc

ALIGNMENTS

RESULT: 11
AAV06369 standard; Protein: 260 AA.

06-SEP-1999 (first entry)

Rhodothermus marinus EGIIT-like cellulase.

Cellulase; endoglucanase; EGIIT; textile; feed additive; baking;

food processing; grain wet milling; pulp; paper.

Rhodothermus marinus.



PF 14-DEC-1998; 98WO-US26552.

PR 16-DEC-1997; 97US-0991720.

PA (GENV) GENECOR INT INC.

PI Bower BS, Fowler T, Phillips JT;

DR WPI: 1999-395187/33.

PT EGIIT like cellulase

PS Example; Fig 6; 47pp; English.

CC The present polypeptide represents a full-length sequence of a

CC novel EgIII-like cellulase isolated from genomic DNA using PCR
CC deduced from a gene sequence isolated from *Rhodothermus marinus*. It was
CC primers (see AA59180-91) based on conserved motifs (see AY06325-29)
CC of *Trichoderma reesei* EgIII cellulase and related enzymes. PCR
CC has been used to identify novel EgIII-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AA0631-70).
CC The sequence shows homology to *T. reesei* EgIII (see AA06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).

Query Match	91.2%	Score 1318.5	DB 20	Length 260
Best Local Similarity	99.3%	Pred. No. 1.7e-118		
Matches 241	Conservative 4	Mismatches 9	Indels 13	Gaps 2

QY	1	MNWKAVLVLSLILLLFGCDMLFPGDNGKEEPEPEPTVELCGRMDARDVAGGRYRVINN	60
Db	1	mnwkvavlvlsllllfgcdwlfpgdgngkepepeptvelcgrmdardvagrgryrinn	60
QY	61	VWGAETAOICIEVGLGETGNFTTRDDHDNGNNVAAPAIIFGCHWGA-----CTSNGSLP	114
Db	61	vwgaetaoicievglgetgnfttrdhdngnnvaapaiifgchwaparaalrdcaragav	120
QY	115	RRVVELSVRFSWLTPTPTTRGRMAAADIMPSPTNGSNGSGGAEIMILNNMGVMPG	174
Db	121	rrvvelsvrfswltptpttrgrmaaadimpsptngsngsggaelimilnnmgvmpg	173
QY	175	GSRVAIVELAGATWEMVYADDMWYIAYRRTPTTSVSELDLKAIFIDAVARGYIREWY	234
Db	174	gsrvatvelagatwemvyadwmwyiayrrtpttsvseldlkafidavargyirpewy	233
QY	235	LHAVETGEIWEKGAGRSADFSFTVQ	261
Db	234	lhavetgeiwegagrltadfsftvq	260

RESULT	2
ABBI4882	
ID	ABBI4882 standard; Protein; 260 AA.
XX	
AC	AABI4882;
XX	
DT	21-NOV-2000 (first entry)
XX	
DE	Emeritcella desertoru EGIII-like cellulase.
XX	
KM	Emeritcella desertoru; Trichoderma reesei; endoglucanase III; EGIII;
KW	cellulase; mutant; enzyme stability; textile treatment;
KW	wood pulp treatment; feed additive; detergent.
XX	
OS	Emeritcella desertoru.
XX	
PN	WO200037614-A2.
XX	
PD	29-JUN-2000.
XX	
PF	12-NOV-1999; 99WO-US26704.
XX	
PR	18-DEC-1998; 98US-0216295.
XX	
PA	(GEMV) GENENCOR INT INC.
XX	
PI	Mitchinson C, Wendt DJ;
XX	
DR	WPI; 2000-482483/42.
XX	
TX	Novel endoglucanase III or endoglucanase III-like cellulase useful for

PT treating textiles and wood pulp comprises a substitution or deletion at
 PN specified positions in the wild form of endoglucanase III -
 XX
 PS Example 1; Fig 3; 52pp; English.

CC The prehn sequence is a cellulase related to endoglucanase III (EGIII)
CC from *Trichoderma reesei*. EGIII-like genes were isolated from genomic DNA
CC libraries constructed from various microorganisms by PCR. The isolated
CC genes showed significant homology to EGIII from *T. reesei*. Certain
CC substitution and deletion mutations have been incorporated into EGIII and
CC EGIII-like cellulases to produce variant enzymes with improved stability
CC e.g., increased resistance to temperature stress. The mutants may be used
CC in textile and wood pulp treatment, as a feed additive, and for reducing
CC biomass to glucose. They are also useful for stonewashing or indigo dyed
CC denim and as an agent in laundry and dish detergents.

SQ Sequence 260 AA;

	Score	DB 21:	Length	260:
Query Match	91.28;			
Best Local Similarity	90.38;			
Best Match 241: Conservative	1.7e-118;			
	4;	Mismatches	9;	Indels
				13;
				Gaps
				2

Qy	1	MNWRVAVLVLSLLILFGDMLPERGDNKKEPEPEPEPVLEICGWMADRDVAGGRVRYINN	60
Dd	1	mmvmavvvisllllfgdcwlfpdagdnkgepepepepvlelcygwadarvagyryvym	60
Qy	61	VWGAETACIEVGLETGNETITRADHDGNVAVYPALYFGCHMGA-----CTSNSGLP	114
Dd	61	vwsgetaqcievgletgnfcltrdhngnvaaypalylfghaparaalrdcaragav	120
Qy	115	RRVQELSVRKSMTLTPTTRGRMAADVIMSPYTNNGNGSGGAEIMILMNNGGMPG	174
Dd	121	rtaheld-----vcpfltcgrnaaqvlwspctnsgngysggaelmlwlnwngympg	173
Qy	175	GSRYATVELAGATWEWIADDMNYIAYRRTPTTSVELDKAFIDDAVARGYIRPEWY	234
Dd	174	gsrvatvelagatwevwadwdmnylaryrtpttsveldkafidavarygylrpewy	233
Qy	235	LHAVETGFELWEGAGLRSAFSTVQ	261
Dd	234	lhavetgfelwegaglrtdafstlvq	260

RESULT 3

ID AAY84347 standard; Protein; 260 AA.

AC AAY84347;

DT 12-JUL-2000 (first entry)

DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.

KW Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;

KW indigo dyed denim; cellulose containing fabric; fabric smoothness;

KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.

OS Rhodothermus marinus.

PN 55 WO200014208-A1

PD 16-MAR-2000

PF 24-AUG-1999; 99WO-US19154.

PR 03-SEP-1998; 98US-0146729.

PA (GEMV) GENENCOR INT INC.

Fowler T;

PI Fowler T;
 XX WPI: 2000-271052/23.
 DR
 XX
 PT Novel variant endoglucanase III-like cellulases with improved
 PT surfactant stability and resistance to temperature stress, useful for
 PT textile processing or cleaning, treating wood pulp, food and grain, and
 PT reducing biomass to glucose
 XX
 XX
 PS Disclosure: Page 65-66, 73pp; English.
 CC The present sequence represents an endoglucanase III (EGIII)-like
 CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
 CC protein. The variant cellulases have improved temperature stability,
 CC and improved surfactant stability. The variant cellulases and
 CC compositions containing them are used in textile processing or cleaning,
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
 CC or appearance of cellulose containing fabrics (e.g. improving fabric
 CC smoothness or removing pills and fibrils). The compositions may also be
 CC used for the removal of immature or dead cotton from cellulosic fibres
 CC or fabric, which can cause uneven dying. The cellulase may also be used
 CC in a detergent composition for washing laundry and dishes and in the
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
 CC The enzymes may also be used in the reduction of biomass to glucose.
 XX
 XX
 S0 Sequence 429 AA;
 Query Match 22.9%; Score 330.5; DB 21; Length 429;
 Best Local Similarity 32.9%; Pred. No. 2.7e-23; Gaps 5;
 Matches 79; Conservative 34; Mismatches 110; Indels 17; Gaps 5;
 QY 24 DGDNGKPEPEPEPTVELCGRMARDYAGGRYRVINNVWGAETACIEVGLTGNTTTR 83
 Db 85 dgsaprtngapksyps-----vfngechytqnmrwgstapqcv-tatdgt-ffvtq 131
 QY 84 ADHNGNNVA--AYPAITFGCHMGACTSNGSLPRVOELSDVRSWLTPTTGRMAAY 141
 Db 132 adgsaprtngapksypsfngechytncspgtldvrltvsapnsysygtfdgagynasy 191
 QY 142 DINESPTVNSGNGYSGAELMTINMNGVMPGSRVATVELAGATWEVYADMDMYIA 201
 Db 192 diwdpdrardg--vngtelmtvfnrvpqlgplspgvtasvgrtvevsggngsndvl 249
 QY 202 YRRTPPTTSVELDLKAFIDDAVARGYTRPEWYLHAVETGFLMEGAGLSADPSVTVQ 261
 Db 250 sfvapsalsgswfdvmfdvratvarglaendwyltsvqagfepwngaglavnsfssive 309
 RESULT 14
 AAW88462
 ID AAW88462 standard; Protein: 261 AA.
 AC AAW88462;
 XX
 XX 10-MAY-1999 (first entry)
 DE Bacillus licheniformis xyloglucanase.
 XX
 XX xyloglucanase; detergent.
 KM
 OS Bacillus licheniformis.
 XX
 FH Key Location/Qualifiers
 FT Protein 30..261
 FT /note="mature protein, this region is specifically
 FT claimed in Claim 19"
 XX
 XX W09902663-A1.
 PN 21-JAN-1999.
 PD
 XX
 PF 01-JUL-1998; 98MO-DK00290.

XX
 PR 24-OCT-1997; 97DK-0001213.
 PR 07-JUL-1997; 97DK-0000822.
 XX
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Bjornvad ME, Jorgensen PL, Outtrup H, Schuelein M;
 DR WPI: 1999-120866/10.
 DR N-PSDB: AAX06949.
 XX
 XX
 PS Claim 29; Page 71-72; 87pp; English.
 CC This polypeptide is an alkaline xyloglucanase obtained from
 CC Bacillus licheniformis ATCC 14580. The enzyme shows optimal
 CC activity at 60 deg C, and retains 50% of its activity after 20 min
 CC at 70 deg. The optimum pH for activity is 5.54, with 32% relative
 CC activity retained at pH 4.59 and 45% at pH 6.49. The coat is
 CC 16.5/sec on xyloglucan at pH 7.5, km 1.1 g/1. The ratio of maximum
 CC xyloglucanase activity to maximum activity on CM-cellulose is at
 CC least 5:1. An isolated polynucleotide (see AAX06949) encoding the
 CC xyloglucanase can be utilised in the production of recombinant
 CC enzymes. Xyloglucanase preparations are useful for improving the
 CC properties of cellulosic fibres, yarn, (non)woven fabric, and for
 CC rating hemp, jute, flax and linen fibres (claimed). They can also
 CC be used in a process for machine treatment of fabrics, and in
 CC detergent compositions (claimed). The enzyme preparations exhibit
 CC high xyloglucanase activity at alkaline pH without essentially
 CC attacking cellulose or cellulose derivatives.
 XX
 S0 Sequence 261 AA;
 Query Match 17.7%; Score 255.5; DB 20; Length 261;
 Best Local Similarity 30.6%; Pred. No. 2.2e-16;
 Matches 70; Conservative 41; Mismatches 85; Indels 33; Gaps 9;
 QY 54 RYRINNVWGAET-----OCIEVGLTGNTTTRADHD-----NGNNVAAYPAITF 100
 Db 45 kyllfnvvgadqyvgwqtl-----yhnsdsdmvwvwnpsntstlvxapsivs 94
 QY 101 GCHW-GACTSNGSLPRVOELSDVRTSMTLTPTTGRMAAYDIFESPVNSGNGYSGA 159
 Db 95 gwhvtegtatgsqfprtlisdqknltkvsysisngcynaaydlwlnhtkcaswdsapt 154
 QY 160 ELMTIWMNGVMPGSGRVATVELAGATWEVW--YAD---WDMNYAYRRTPPTTSVSE 213
 Db 155 eimtlwn-ntnagpagsvvetvsigshwkykylidagggkvwvfwfslrtantgs-an 212
 QY 214 LDLKAFIDD-AVARGYTRPEWYLHAVETGFLMEGAGLSADPSVTVQ 261
 Db 213 Inrdftnyladsqwkstkyssvefgevfgtqglinswdcvr 261
 RESULT 15
 AAY06370
 ID AAY06370 standard; Protein: 264 AA.
 AC AAY06370;
 XX
 XX 06-SEP-1999 (first entry)
 DE Erwinia carotovora EGIII-like cellulase.
 XX
 XX Cellulase; endoglucanase; EGIII; textile; feed additive; baking;
 KM food processing; grain wet milling; pulp; paper.
 XX
 OS Erwinia carotovora.


```
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. 6268328 Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/321,981
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRF
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" = Streptomyces
; OTHER INFORMATION: thermovibaceus
US-09-321-981-5
```

```
Query Match 27.7%; Score 400.5; DB 4; Length 386;
Best Local Similarity 39.0%; Pred. No. 5,8e-32;
Matches 90; Conservative 28; Mismatches 104; Indels 9; Gaps 6;
```

```
OY 33 PEPPVELCGRMDARVAGRYVINNVWGAETACIEVLEGTGN-FTTRADHDGNN 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 42 PPAQANQICDRGTITIQ-DRYVONNRRGTSATOCINV---TGNGFELTQADGSPVTN 97
OY 92 VA--AVALPVGCHMGACTNSGLPRRVOELSDVRSWTLPTTGRWNAAYDIWFSPTV 149
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 GAPKSYSPYDGCYHCACARLTLPKRISIGSAPSSVSRYTCNGVYNAAYDIWLDLP-T 156
OY 150 NSGNGYSGGELMILWLNMGVMPGSGRVATVELAGATWEVYADWDMNTIARRTTPTT 209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 157 PRNGVAV-REIMTFNRVGPVOPISPGVTAHVGSRMEVWTCGNSGNDVIFSLAPSAI 215
OY 210 SVSELDKATIDAVANGYIRPEWYLAHVEFGFELMEGAGLSADGSPVTN 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 SSMSFDVQKVDQAVSHGLATPDWYLTISIQAGFEPMEGGLAVNSTRSSAV 266
```

```
RESULT 3
US-09-216-295-21
; Sequence 21, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328 Variant Egitr-like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 312
; TYPE: PRF
; ORGANISM: Actinomycete 11AG8
US-09-216-295-21
```

```
Query Match 27.4%; Score 396.5; DB 4; Length 312;
Best Local Similarity 39.7%; Pred. No. 1.1e-31;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;
```

```
OY 40 ELGCRMDARVAGRYVINNVWGAETACIEVLEGTGN-FTTRADHDGNNVA--AYP 96
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 QICDRGTITIQ-DRYVONNRRGTSATOCINV---TGNGFELTQADGSPVTN 89
```

```
OY 97 AIYFGCHMGACTNSGLPRRVOELSDVRSWTLPTTGRWNAAYDIWSPVTNSGNGYS 156
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 SYVDGCHYGACARLTLPKRISIGSAPSSVSRYTCNGVYNAAYDIWLDLP-TPRTNGVN 148
OY 157 GGAELMTWLNMGVMPGSGRVATVELAGATWEVYADWDMNTIARRTTPTTSVELDL 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 -REIMTFNRVGPVOPISPGVTAHVGSRMEVWTCGNSGNDVIFSLAPSAISSFDV 207
OY 217 KAFIDAVANGYIRPEWYLAHVEFGFELMEGAGLSADGSPVTN 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 KDEVDQAVSHGLATPDWYLTISIQAGFEPMEGGLAVNSTRSSAV 251
```

```
RESULT 4
US-09-104-308-1
; Sequence 1, Application US/09104308
; Patent No. 6187577
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: Van Der Kleij, Wilhelmus A.H.
; APPLICANT: Van Solingen, Piet
; APPLICANT: Meyler, Walter
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,308
; FILING DATE: 24-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/974,042
; FILING DATE: 19-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC539
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7555
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-104-308-1
```

```
Query Match 27.4%; Score 396.5; DB 4; Length 371;
Best Local Similarity 39.7%; Pred. No. 1.4e-31;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;
```

```
OY 40 ELGCRMDARVAGRYVINNVWGAETACIEVLEGTGN-FTTRADHDGNNVA--AYP 96
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 QICDRGTITIQ-DRYVONNRRGTSATOCINV---TGNGFELTQADGSPVTN 89
OY 97 AIYFGCHMGACTNSGLPRRVOELSDVRSWTLPTTGRWNAAYDIWSPVTNSGNGYS 156
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 SYVDGCHYGACARLTLPKRISIGSAPSSVSRYTCNGVYNAAYDIWLDLP-TPRTNGVN 148
OY 157 GGAELMTWLNMGVMPGSGRVATVELAGATWEVYADWDMNTIARRTTPTTSVELDL 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```

DB      213  LNIHFRTNVVLADSKQWMLSKTKYVSSVEGTETVEFGGTGGINISNMDDVTNR 261

RESULT      8
US-08-032-848C-12
; Sequence 12, Application US/08032848C
; Patent No. 5475101
; GENERAL INFORMATION:
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen A.
; APPLICANT: Weiss, Geoffrey L.
; APPLICANT: Larenas, Edward
; APPLICANT: Lorch, Jeffrey D.
; TITLE OF INVENTION: Purification and Molecular Cloning of
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/032,848C
; FILING DATE: MAR 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 742-7356
; TELEFAX: 415 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 233 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-032-848C-12

Query Match      17.6%; Score 254; DB 1; Length 233;
Best Local Similarity 32.0%; Pred No. 1.2e-17;
Matches 73; Conservative 39; Mismatches 86; Indels 30; Gaps 10.

OY      54  RYRVINNWGAETAQ-----CIEVGLTGNFTTRADHDNGNNVAAYPAIFYGCH 103
      : : | | | | : : : : | : | : : : | | | : : | : | : |
Db      16  KYLLFNWVWGKDEIKGMQOTIFYNPSISMG---MWMHMPSSYH---SVKAYPSLVSGNH 68

OY      104 WCA-CTSNSGCLPRRQELSDVRTSMTLPRITTGRRNNAAYDIFFSYTNSGNGYSGAGELM 162
      : : | | | | | : : : : : | : | | | | | : : | | | | |
Db      69  WTAGTENSGLDITQSSNKSITSNVTYSIKATGYNAAYDIWHTTDXANWDSPTDELH 128

OY      163 IWLN-WNGGVMPGSGSRVATVELAGATWEY---WYADWD---WNYIAYRRTPTTSVSEL 214
      | | | | | : : : : | : | : : | : | : : | : | : : | : |
Db      129 IWLNDTNG--PAGDYIEIVFLGDSSWMYFKWMINNADNGGMNFFSVHTSGTNSAS-L 185

OY      215 DKAATIDAV-ARGITRPEWYLHAAVETGFELMEGAGLRSDAFSVYQ 261
      : : | : | : | : : : : : | : | : | : : : : : : : : :
Db      186 NIRHFTDVLVOTKQWMSDEKYYISVEFGETIFGGGQIDITFEMRVYK 233

```

```
; GENERAL INFORMATION:  
; APPLICANT: Mitchinson, Colin  
; APPLICANT: Mendt, Dan J.  
; TITLE OF INVENTION: No. 6268329e1 Variant EGIII-Like Cellulase Compositions  
; FILE REFERENCE: GC555  
; CURRENT APPLICATION NUMBER: US/09/216_295  
; CURRENT FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 24  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Erwinia carotovora  
US-09-216-295-24  
  
Query Match      17.0%; Score 246; DB 4; Length 263;  
Best Local Similarity 31.6%; Pred. No. 8,6e-17;  
Matches 72; Conservative 40; Mismatches 84; Indels 32; Gaps 11;  
  
OY    54 RRYVNNWGAETAO-----CLEVGLETGFTTRADHDNGNNAAYPAIFECF 103  
Db     :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
48 KYLLFNWNWGKDEIGMOOTTFNSPISMG---NMHWPSSYH----SVKAPPSLVSGWH 100  
  
OY    104 WCA-CTNSGLPRRQVELSDVRTSWTLPIITGRBNAAYDIWFESPVTNSGNGSCGAEIM 162  
Db     |||::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
101 WTAGTTENSGLEIQLSSNKSTITSNTYSIKATGIYNAAIDWEHTTDXANMDSPTDEL 160  
  
OY    163 IWL-N-WNGVMPGGSRVAIVELAGATWEWYAWD-----DMNYIAVRRTPTTSSEL 214  
Db     |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
161 IWLNDTNMG--PADGYIEFLFGDSMWV-FKGMINADNGGMN-VSPVHTSGTMSAS-L 215  
  
OY    215 DKAATIDAAY-AKGIREDWYLHAIVTEGFELMEGAGLRSADEFSTVO 261  
Db     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
216 NRRHPTDYLVOTKWMSDEKYISSVEFEITFGDGQIDITEWRVDVK 263  
  
RESULT 10  
US-09-216-295-5  
; Sequence 5, Application US/09216295  
; Patent No. 6268328  
; GENERAL INFORMATION:  
; APPLICANT: Mitchinson, Colin  
; APPLICANT: Mendt, Dan J.  
; TITLE OF INVENTION: No. 6268329e1 Variant EGIII-Like Cellulase Compositions  
; FILE REFERENCE: GC555  
; CURRENT APPLICATION NUMBER: US/09/216_295  
; CURRENT FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 259  
; TYPE: PRT  
; ORGANISM: Aspergillus aculeatus  
US-09-216-295-5  
  
Query Match      15.5%; Score 224, 5; DB 4; Length 259;  
Best Local Similarity 29.6%; Pred. No. 1,2e-18;  
Matches 76; Conservative 36; Mismatches 92; Indels 53; Gaps 12;  
  
OY    40 ELICGWMDARDVAGRGYYRIYNNWGAET--AQCIEV--GLETGNETITRADHDNG-NNVA 93  
Db     :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
20 QUCDDY--ATTYGVIYTINNINLMGDAGSGCCTTVNSASSAGTSWSRKMMMSGGENSVK 77  
  
OY    94 AYPALPYFCGHWCATSNSGL---PRRQVELSDVRTSWTLPIITPG-RKNAAVIDWFEPVT 149  
Db     :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
78 SY-----ANSGLTFKKLYAQISOIPTTARMSYDNMGIRADVAVDLETTADI 124  
  
OY    150 NSGNYSGGAGELMTLANNNGVMPCGSRVATVELAGATWEWY-ADMWMNIIAARRTPPT 208  
Db     :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
125 NHVT-WSGDYEELMTILARYGVOPIGSQIATPVLDGOWTELWGANGSOAKTYSEVPAPPI 183
```



```

OY      209  TSVSELDKAFID-----DAVARGYIR-----DEWYLHAETGEE 243
Db      184  TSF-QGDVNDFFKXITQNHGCFPASSQYLITLQFGTEPFTGCPATLSVNSMSASVQOAGFE 242
OY      244  LMEGCAGLRADPSVTV 260
Db      243  PWONAGGLAVNSFSSTV 259

RESULT 11
US-09-216-295-7
; Sequence 7, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EGIIT-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aspergillus kawachii (2)
US-09-216-295-7

Query Match 14.8%; Score 214; DB 4; Length 239;
Best Local Similarity 32.1%; Pred. No. 1,2e-13;
Matches 78; Conservative 37; Mismatches 84; Indels 44; Gaps 15;

OY      40  ELGCGMDARDVAGGKRXRYINWVGAET--AGCIEY--GLETGNFTTRADHDNG-NNVA 93
Db      20  QLCQDY--ATYTGYYTTINNMLMGKADAGSGQCTTVNSASSAGTSWSTKMWNSGGENSVK 77
OY      94  AYPALIEGCHMGACTSNSSL--PRVOELSDV---RTSWLPIPTTGRNNAAYDIWFS 146
Db      78  SY-----ANGLSFPNKKLVQSQISHIIPMAAMWSINDNICIRGR--AYDLFTA 121
OY      147  PVTNSGNGYSGCAELMIWLNNGGVPGPSRAVATVELAGATVEWY---ADMDWNYIAY 202
Db      122  ADINHVY--WSDGYELMIWLARGVQVPLGSOIATVTEGQWTWELWYGVNCAQKTSFYA- 179
OY      203  RRTPTTSVSELDKAFIDDAVA--RGYIRPEYVL--HAYVGFLMEGCG-AGLRADPSV 258
Db      180  --ANDPISY--QGDINDFFKXILTQNHGCFPASSQYLITLILALQGTETPFTGCPATLVNADWSA 236
OY      259  TVQ 261
Db      237  SVQ 239

RESULT 12
US-09-216-295-10
; Sequence 10, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EGIIT-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Humicola insolens
US-09-216-295-10

```

	Query Match	14.8%	Score 213.5	DB 4	Length 253	
	Best Local Similarity	30.5%	Pred. 0.1-4e-13			
	Matches	73	Conservative	35	Mismatches	98
					Indels	33
					Gaps	13
QY	34	EPEPTVELC---	GRMDARDVAGGRYRVINNWGAETA---	OCIEV-GLETFGNFTITRAD	85	
Db	28	EPROIRSLCELYGYV----	SGNGYELLNNLMKGTATSGMOCYTLDDTNNNGIOWSTAW		82	
QY	86	HDNG--NNVAAYPATYFCGCHWGCACTNSGGLPRVVELSDVPT--	SWTLPLPTTGGWNAV	141		
Db	83	EMOGAPDNVKSFPYVGKRIQRC-----	RKIDISMTSPYSWTIDR-TDIRANAY	132		
QY	142	DWESPVTVNSGNGYSGAELMIIMLNWNGGVMPGSGRVAIVELAGATWEWVADMDWNYIA	201			
Db	133	DVETARDDPHN-MGGDDIELMIIMILARYGVIYPIGTFHGOVNLAGRITWMLD--	NGYGNRNVR	190		
QY	202	YRRTPPTSYSGL--DLKAFITDDANA--RGYLRPEYVLAHVEGFLMEGG--AGLSADPF	256			
Db	191	YSLFPESSDIDRFSCDIDGFENYLERNNHGYPAREONLIYVQGTCEFGGPRPRTCRDF	249			

```

RESULT 13
US-08-032-848C-13
; Sequence 13, Application US/08032848C
; Patent No. 5475101
GENERAL INFORMATION:
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Weiss, Geoffrey L.
APPLICANT: Larena, Edward
APPLICANT: Lorch, Jeffrey D.
TITLE OF INVENTION: Purification and Molecular Cloning of
TITLE OF INVENTION: EG III Cellulase
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/032.848C
FILING DATE: MAR 17 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 742-7356
TELEFAX: 415 742-7217
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-032-848C-13

Query Match          14.7%; Score 213; DB 1; Length 221;
Best Local Similarity 31.1%; Pred. No. 1,3e-13;
Matches    73; Conservative   40; Mismatches   92; Indels    30; Gaps     12;

```

```

Db      4 QLCDXY--ATYGGVYTTNNLMGKDGSGSQCCTVNSASSAGTSMWTKMWSGGENSVK 61
OY      94 AYPATFGCHMGACTNSGL---PRRQVELSDVRTSWTLPIITG--RWNAAVDIMWSPVT 149
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      62 SY-----ANSGLTFENKLLVSQISQIPTTARMSYDNTGIRADVAITDLFTRADI 108
OY      150 NSNGYSGAELMIMLMNMGVMPGSSRVATVELAGATWEVWY-ADWDNNTIAYRRTTPT 208
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      109 NHVT-NSGDYELMIMLARVGGVQPIGQIATATVDGQWELMWGANGSQKTSFVAPTPI 167
OY      209 TSVSELDKAFIDDAVA-RGYRPEWYLHAVETGELMEG-AGLRADSVTYQ 261
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
        168 TSE-QGDVNDFFRYLQNHGFPASSOYLITLQCTEPTGTPATLVSNSASVQ 221

RESULT 14
US-09-216-295-20
; Sequence 20, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Emeritella desertoru
US-09-216-295-20

Query Match      14.6%; Score 210.5; DB 4; Length 246;
Best Local Similarity 28.1%; Pred. No. 2.7e-13;
Matches 71; Conservative 33; Mismatches 80; Indels 69; Gaps 12;

OY      40 ELGRDARDVAGGRYRVINNWAETACIEVLEGTNFTTRADHDNGNNVAAYPAIR 99
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      30 DFCGQMDTATV--GNFIVYNNLMGQDNA---DSGSQTG-----VDSANGNSIS----- 72
OY      100 FCGHMGACTNSG-----LPRRQVELSDVRTSWTLPIITGRW----- 137
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      73 ---WHTTWSMSSGSSSVKSYANAAVOFTSKLNSISPTSW-----KQVSTTDI 120
OY      138 --NAAVDIMWSPYTNNGNGYSGAELMIMLMNMGVMP---GSSRVATVELAGATWEVWY 192
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      121 VANVAITDLF---TSSAGGDSEYELMIMLARVGGVQPIGQIATATVDGQWELMWGANG 176
OY      193 AD---MDWNNTIAYRRTTPTTTSVSELDKAFIDDAVARGYRPEWYLHAVETGELMEG- 247
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      177 GPRGSMQVYSFVA---SSPTSESFADLMDFIYVLAENGLSSQYLYTHQAGTEPPTGT 232
OY      248 GAGLRADSVTYQ 260
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      233 DATLTVSSYSYSV 245

RESULT 15
US-09-216-295-9
; Sequence 9, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 9
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Humicola griset
US-09-216-295-9

```

```

Query Match      14.6%; Score 210.5; DB 4; Length 253;
Best Local Similarity 28.2%; Pred. No. 2.8e-13;
Matches 70; Conservative 31; Mismatches 96; Indels 51; Gaps 12;

OY      34 EPEPTVELC---GRMDARDVAGGRYRVINNWAETACIEVLEGTNFTTRADHDNGN 90
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      28 EPRQISLCELXYGW-----SGNGYEILNLMGKDIA-----TSGMCCTYLDEGTNNG 74
OY      91 NVAAYPAIRYFGCHMGACTNSGLP-----RRRQVELSDVRT---SWTLTPI 132
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      75 -----GIQWNTAMEWQAPDVKNYPIYGKQIQRGKISDINSMTSVSWTYDR- 123
OY      133 TTRMNAAYDIMFSPYTNNGNGYSGAELMIMLMNMGVMPGSSRVATVELAGATWEVWY 192
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      124 TDLRANAYVDVFTARDDPHN-WGQYELMIMLARVGGVQPIGTPHSQVVLAGRITDLM- 181
OY      193 ADWDNNTIAYRRTTPTTTSVSEL--DKAFITDDAVA-RGYRPEWYLHAVETGELMEG- 248
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      182 TGYNGNMRYVSELPSPGDIRDFSCDIKDFPNYLERNHGYPARQONLIYVOVTECFETGCP 241
OY      249 AGLRADF 256
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      242 ARFTCRDF 249

```

Search completed: August 15, 2002, 10:52:08
 Job time: 172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:48:41 ; Search time 29.16 Seconds

(without alignments)
860.059 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446

Sequence: 1 MNVRAVLVLLLLFGCDM.....PELMGAGLRSADESVTVQ 261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	352	24.3	382	2	JC2571
2	252.5	17.5	264	2	JU0328
3	213	14.7	237	2	S12610
4	202	14.0	274	2	A72241
5	194	13.4	151	2	H70895
6	163.5	11.3	239	2	S55931
7	153	10.6	258	2	H72240
8	142.5	9.9	332	2	G90291
9	139	9.6	334	2	G90360
10	99	6.8	957	2	H82261
11	97	6.7	673	2	H70528
12	95.5	6.6	322	2	H90425
13	94	6.5	443	2	D72204
14	93	6.4	604	2	B84221
15	93	6.4	741	2	G69514
16	91.5	6.3	690	2	T27357
17	91	6.3	950	2	H87611
18	88.5	6.1	261	1	S12745
19	87.5	6.1	317	2	T36926
20	87	6.0	286	2	S48201
21	87	6.0	1449	1	VG1HFS
22	86.5	6.0	596	2	A55976
23	85	5.9	475	2	T35697
24	85	5.9	497	2	H83886
25	84.5	5.8	198	2	AH2306
26	84.5	5.8	610	2	JH0573
27	84.5	5.8	765	2	S76795
28	84	5.8	468	2	AD0585
29	84	5.8	656	2	H84649

ALIGNMENTS

30	83	5.7	263	2	S06330	rlcin E - castor b
31	83	5.7	302	2	H90717	hypothetical prote
32	83	5.7	323	2	H85567	hypothetical prote
33	83	5.7	468	2	H64802	ybfm protein - Esc
34	83	5.7	574	2	F75356	serine/threonine p
35	83	5.7	595	2	T29434	beta-galactosidase
36	82.5	5.7	1090	2	A40653	chitinase (EC 3.2.
37	82.5	5.7	1449	2	S59077	cellulose 1,4-beta
38	82.5	5.7	1449	2	A43573	E2 glycoprotein pr
39	82	5.7	576	1	RUCSD	rlcin D precursor
40	81.5	5.6	544	2	A47726	dis1-suppressing p
41	81.5	5.6	1374	2	AE3259	extracellular seri
42	81.5	5.6	1447	1	VG1HE2	E2 glycoprotein pr
43	81.5	5.6	1447	1	VG1HE3	E2 glycoprotein pr
44	81	5.6	474	2	G72658	probable Vir B11 A
45	81	5.6	3670	2	T36249	CDA peptide synthe

RESULT 1

JC2571 cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)

N/Alternate names: endo-1,4-beta-glucanase; endoglucanase

C/Date: 13-Jun-1995 #sequence, revision 14-Jul-1995 #text, change 22-Oct-1999

C/Accession: JC2571; S34392

RefSeq: B1: Hanhart, E.; Irdani, T.; Egbal, M.; McCarthy, A.J.; Mastromeli, G.

Gene 148, 119-124, 1994

A/Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase

A/Reference number: JC2571, MUID:95011642

A/Accession: JC2571

A/Molecule type: DNA

A/Residues: 1-382 <PER>

A/Cross-references: EMBL:X7953; NID:9393391; PIDN:CAA52139.1; PID:9393392

A/Note: this cellulolytic strain was isolated from the gut of termites

C/Genetics:

A/Function:

A/Description:

A/Pathway:

C/Superfamily:

C/Keywords:

F:1-37/Domain:

F:38-382/Product:

F:279-380/Domain:

F:280-379/Dissulfide bonds:

#status predicted

#status predicted

#status predicted

#status predicted

#status predicted

OY 250 -----GLRSADFSTV 260
 Db 255 NTAAKFWTFRDFSEV 272

RESULT 5

hypothetical protein Rv1090 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70895

R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MID:98295987

A:Accession: H70895

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:AL021897; GB:AL123456; NID:g3256022; PIDN:CAAL7206.1; PID:e125196

A:Experimental source: strain H37RV

C:Genetics:

Query Match 13.4%; Score 194; DB 2; Length 151;
 Best Local Similarity 33.3%; Pred. No. 4.7e-09;
 Matches 51; Conservative 16; Mismatches 78; Indels 8; Gaps 4;

OY 111 GSPRRVQELSDVTSMTLPTTGRNNAVDIWF--SPVTNSGNGYSGAELMIMLNM 168
 Db 3 TNPTEVGOILSAPTSIDYNYPTTGWDASDYDCLDSTPKTGYN---QOEIMTFNHQ 58

OY 169 GGMPPGSRVATVELAGATWEVYADMDM-NYIAYRTPTTSVELDKAFIDDAVARG 227
 Db 59 GSQIPVGSPPGNTTTEKKNVMDGSGNMMNAAYATPE-IEWSEFVMSFVDHATME 117

OY 228 YIRPEWYLAHVETGFELMEGAGLRSADEFSTV 260
 Db 118 PTDMSWLTSTIRAGLEPWSGCVGLGVDSEFAKV 150

RESULT 6

cellulase (EC 3.2.1.4) precursor - Aspergillus niger

N:Alternate names: carboxymethylcellulase I; endo-1,4-beta-glucanase

C:Species: Aspergillus niger
 C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: S55931; S60657; J50730

R:Sakamoto, S.; Tamura, G.; Ito, K.; Ishikawa, T.; Iwano, K.; Nishiya, N.

Curr. Genet. 27, 435-439, 1995

A>Title: Cloning and sequencing of cellulase cDNA from Aspergillus kawachii and its expression

A:Reference number: S55931; MID:96059347

A:Accession: S55931

A:Molecule type: mRNA

A:Residues: 1-239 <SAK>

A:Cross-references: EMBL:D12901; NID:g217812; PID:g217813

A>Note: the source is designated as Aspergillus kawachii

A:Molecule type: protein

A:Residues: 76-86;176-186 <SAK2>

A>Note: the source is designated as Aspergillus kawachii

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose, cellulose degradation

C:Keywords: blocked amino end; glucosidase; hydrolase; polysaccharide degradation

F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-239/Product: cellulase #status predicted <MAT>

Query Match 11.3%; Score 163.5; DB 2; Length 239;
 Best Local Similarity 27.3%; Pred. No. 2.7e-06;
 Matches 63; Conservative 38; Mismatches 97; Indels 33; Gaps 11;

OY 41 LCGRMARDVAGGRVYINNVWG---AETACIEVG--LETGFTTRADHDNG-NNVAA 94
 Db 19 MCSQYDS--ASSPPYSVNOMLWGEYGTGQCYVYVKLSSGASWHKWTWSGEGTVKS 76

OY 95 YPAIVFGCHWAGTSSNGL---PRVQELSDVTSMTLPTTGG-RMNAYDIWFSPTVN 150
 Db 77 Y-----SNSGLFDKLVSDVSIPTSTWSDDDINQADVSYDL-FTANA 122

OY 151 SGNGYSGAELMIMLWNGGVMPPGSRVATVELAGATWEVY----ADMDNYIAYRRT 205
 Db 123 DHATSSGDDELMIWILARVGSVQPIKQIATATVGGKSEWYGTSTQACAEQRTYSFVAG 182

OY 206 TPTTSVELDKAFIDDAV-RGYIRPEWYLAHVETGFELMEGAGLRSD 255
 Db 183 SPINSMG-DIKDFNLTQNGCPASSQHLITLQGTETPTGPAFTYVD 232

RESULT 7

endoglucanase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: H72240

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

C.M. Nature 399, 323-329, 1999

A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MID:99287316

A:Accession: H72240

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <ARN>

A:Cross-references: GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AAD36591.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

Query Match 10.6%; Score 153; DB 2; Length 258;
 Best Local Similarity 27.7%; Pred. No. 2.1e-05;
 Matches 51; Conservative 24; Mismatches 87; Indels 22; Gaps 8;

OY 92 VAAVPAIVFGCH-WGACTS-NSGLPRVQELS--DVRTSWTLPTTGRNNAVDIWFSP 147
 Db 63 VLGTPERYIGKRWENHTAGSKLPVPVSSKSFSEVSDIHHERSLPLNFMETWLTTR 122

OY 148 VTNSGNGYSGAELMIMLWNGGVMPPGSR-----VATVELAGATWEVYADMDNY 199
 Db 123 EKQVTEASIDDELWMEYFN-NLTGCGEKIEFTTIPFVLNGSSVECTWELMAENGMDY 181

OY 200 IAYRRTPPTTSVS-EIDDKAFIDDA-----VARGIIRPEWYLAHVETGFELMEGAGLR 252
 Db 182 LAFRLDPPVKRGRKEDVRFLDPAAGKALSSSARVXDDEFDLFTWEIGTER--GSPETK 239

OY 253 SADF 256
 Db 240 SAQF 243

RESULT 8

endoglucanase precursor [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: G90291

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.


```

Db 107 VARAPVAVGGLPGVAVLF-----RGRDDDEVLALAGLAFPTLVYYSRFRMSDYLVA 162
OY 54 RYRVINNVWGAETAGCIEVGETETGFTTRADHDNG-----NNVAA 94
Db 163 AF-----SLAAGFAV-RA-HDTGRRLLPVAGMLALATLAKENALV 203
OY 95 YPAIYEGCHWGACTNSGLPRVOELSDVRSWTLPITTGKRNAAVDIMFSPVYNSGNG 154
Db 204 YAMFAGAG--GALVAD-----RLLITANPRLGSLWTS--LQAGVTRRAARGFAMRRTTLASA 256
OY 155 YSGGAELMTWLNKNN--GCY-----MPGGSRAVVELAGATWEVYADW---DMNTIAY 202
Db 257 VVAGAVFAVFPYPRPVAGGIGAAPTRLPSAVAGSADAHAALMGTWYVDGVDKDHSTIAY 316
OY 203 ----RRTPTTSV 211
Db 317 LVVALRTLATTAIV 329

```

RESULT 15

G69514

hypothetical protein AF2119 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Nov-1999
 C:Accession: G69514

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 Glodex, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

A:Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaea
 A:Reference number: A69250; M01D:98049343

A:Accession: G69514
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-741 <RLE>
 A:Cross-references: GB:AE000958; GB:AE000782; NID:92689281; PIDN:AA89151.1; PID:9264842

C:superfamily: Archaeoglobus fulgidus hypothetical protein AF2119

Query Match

Best Local Similarity 24.4%; Score 93; DB 2; Length 741;
 Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

```

OY 75 ETGNFTITRADHDNGNNVAYPA-----IYRGCHWGACTNSG-----LPRR-- 116
Db 204 DTDITTLVAKVYNSGNPNVDYVKFAEPDSQRIFLG---GAMTNSGIAKLSFIPKNVG 260
OY 117 -----VOELSDVRSWTLPITTGKRNAAVDIMFSP---VTNSGN-GYSGGAELM 162
Db 261 LSDKLRVNFVAKIEDVMTNCNAYTTTNRAILAEVVAITPGSYDITLVGRMYTSGGAD-W 319
OY 163 IWLNV-----NGVMP-----GGSRAVVELAGATWEVY----- 192
Db 320 VRVWVYVDENVSLKPIPVYQKFTGNR-----ASVTIMKYGLDDYCTDPNCHREGI 371
OY 193 -----ADMGMNTIAYRRTTP 207
Db 372 YGNFDADWDGACIAGVSTTP 392

```

Search completed: August 15, 2002, 10:51:39
 Job time: 178 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:51:42 ; Search time 15.93 Seconds

(without alignments)
634.388 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446
Sequence: 1 MNMRAVLVLSLLIFGCDW.....FELMEGAGLSRADEFSVTQ 261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	252.5	17.5	264	1 GUNS_ERMCA	P16630 erwina car
2	213	14.7	237	1 GUN_ASPAC	P22659 aspergillus
3	163.5	11.3	239	1 GUN_ASPAC	Q12679 aspergillus
4	93	6.4	625	1 BGAL_LACSK	Q48846 lactobacill
5	93	6.4	741	1 YL19_ARCFU	Q28161 archaeoglob
6	88.5	6.1	261	1 XYNA_CLOSA	P17137 clostridium
7	88	6.1	269	1 EL2_BOVIN	Q29461 bos taurus
8	87	6.0	286	1 GUB_RHOHR	P18450 rhodothermu
9	87	6.0	1449	1 VGL2_CVPEF	054161 porcine tra
10	85	5.9	475	1 ABFB_STRCO	P11220 streptomyce
11	84.5	5.8	610	1 CHIT_STRPL	P02167 porcine tra
12	84	5.8	1447	1 VGL2_CVPEF	P75757 escherichia
13	83	5.7	468	1 YBFW_ECOLI	P96463 streptomyce
14	83	5.7	475	1 ABFB_STRPL	P50899 cellulomona
15	82.5	5.7	1090	1 GUXB_CELFI	P33470 porcine tra
16	82.5	5.7	1449	1 VGL2_CVPEF	P02879 rictinus com
17	82	5.7	576	1 RIC1_RICCO	O9CK15 pasteurella
18	81.5	5.6	224	1 GPH_PASMU	P71121 corynebacte
19	81.5	5.6	225	1 THTR_CORGL	P36616 schizosacch
20	81.5	5.6	544	1 DSK1_SCHPO	P07946 porcine tra
21	81.5	5.6	1447	1 VGL2_CVPEF	P32762 streptococ
22	80.5	5.6	318	1 ALYS_BPHB3	O10534 mycobacteri
23	80.5	5.6	376	1 SERC_MYCTU	P21817 homo sapien
24	80	5.5	5038	1 RYR1_HUMAN	P76115 escherichia
25	79.5	5.5	700	1 YNCD_ECOLI	P32823 altermonas
26	79.5	5.5	820	1 CHIA_ALTSO	P23659 clostridium
27	79.5	5.5	986	1 GUNZ_CLOSR	Q07833 bacillus su
28	79	5.5	2334	1 WAPA_BACSU	P55335 magnaporthe
29	78.5	5.4	233	1 XYNA_MAGGR	P58029 sulfobolus
30	78.5	5.4	479	1 CBSA_SULSO	P1626 caprine art
31	78.5	5.4	966	1 ENV_CAEYC	O01977 porcine tra
32	78.5	5.4	1447	1 VGL2_CVPEF	Q9ut44 schizosacch
33	78	5.4	509	1 YAGE_SCHPO	

34	78	5.4	852	1 POL_BLVJ	P03361 bovine leuk
35	77.5	5.4	822	1 ACM1_DROME	P16395 drosophila
36	77.5	5.4	746	1 PAC_ECOLI	P06875 escherichia
37	77.5	5.4	877	1 AGUL_HORVU	Q43763 hordeum vul
38	77.5	5.4	1356	1 HET1_PODAN	Q00808 podospora a
39	77	5.3	572	1 SYM_AERPE	Q9cy33 aeropyrum p
40	77	5.3	1481	1 APV1_THEET	P38399 t amylopull
41	77	5.3	2044	1 SIF2_DROME	P1620 drosophila
42	76.5	5.3	216	1 SPRB_IPOBA	P10965 ipomoea bat
43	76.5	5.3	270	1 KITM_MOUSE	Q9I088 mus musculu
44	76.5	5.3	318	1 ALYS_STRPN	P06653 streptococ
45	76.5	5.3	327	1 A85B_MYCLE	P31951 mycobacteri

ALIGNMENTS

RESULT 1	ID	GUNS_ERMCA	STANDARD	PRT	264 AA
AC	P16630				
DT	01-AUG-1990 (Rel. 15, Created)				
DT	01-AUG-1990 (Rel. 15, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Endoglucanase S precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase S) (cellulase S).				
GN	CELS.				
OS	Erwinia carotovora.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Pectobacterium.				
OX	NCBI_TaxID=554;				
RN	(1)				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=SCC3193;				
RA	MEDLINE=90337352; PubMed=2379837;				
RA	Saarialhti H.T., Henttinen B., Palva E.T.;				
RT	"Cels": a novel endoglucanase identified from Erwinia carotovora				
RT	subsp. carotovora."				
RL	Gene 90:9-14 (1990).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL				
CC	HYDROLASES).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL: M33399; AAA24817.1; -				
DR	PIR: J00328; J00328.				
DR	InterPro: IPR002594; Glyco_hydro_12.				
DR	Pfam: PF01670; Glyco_hydro_12; 1.				
DR	ProDom: PD004316; Glyco_hydro_12; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
FT	SIGNAL 1 32				
FT	CHAIN 33 264				
FT	SEQUENCE 264 AA: 29757 MW: E6D61388950C77AA CRC64;				

Query Match 17.5%; Score 252.5; DB 1; Length 264;
Best Local Similarity 31.6%; Pred. No. 1,4e-14;
Matches 72; Conservative 39; Mismatches 86; Indels 31; Gaps 10;

QY	54	RYRYINNVGAETIAQ-----CIEVGLTEGNFTITRADHDGNNVAAYPAITFGCH	103
DB	48	KYVLENNVWGKDEIKGQGITFYNSPISMG---WNHWPSTH---SVKAYPSLVSGWH	100
OY	104	WGA-CTNSGGLPRVQGLSVYRSTWTLPTITGWMNAYDIWFSPVTSNGSGAGELM	162

```

Db 101 WTAGTENSGLPIOLSSKSTTSNTVYSIKATGYNAYDIWFTTDKANNSSPTDELM 160
OY 163 TWLN-WNGGVWPGSSRAVTEVLGATWEVWADM-----DNNYIAYRRTPTTSVSEL 214
Db 161 IWLNDTNAG--PAGDYIETVELGDSNNV--FKGWINADNGGNVVFVHTSGNSAS-L 216
OY 215 DLKAFIDAV-ARGYIRPEWYLAHVETGFEIMEGACLRSDSVYQ 261
Db 217 NHRFTDYLVTOKOMSDERTISSVEFGTEIFGDDQIDITEMRVDK 264

RESULT 2
GUN_ASPAC
ID GUN_ASPAC STANDARD: PRT: 237 AA.
AC P22669:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase) (FI-CMCase).
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F-50;
RX MEDLINE=91016934; PubMed=2216782;
RA Ooi T., Shimmyo A., Okada H., Murao S., Kawaguchi T., Arai M.;
RT "Complete nucleotide sequence of a gene coding for Aspergillus
RL aculeatus cellulase (FI-CMCase).";
RL Nucleic Acids Res. 18:5884-5884(1990).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=F-50;
RX MEDLINE=91064758; PubMed=2249253;
RA Ooi T., Shimmyo A., Okada H., Hara S., Ikenaka T., Murao S.,
RA Arai M.;
RT "Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase)
RL from Aspergillus aculeatus.";
RL Curr. Genet. 18:217-222(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- INDUCTION: BY CELLULOSIC MATERIALS AND HEMICELLULOSES.
CC -1- MISCELLANEOUS: WILK ALSO HYDROLYSE 1,4-LINKAGES IN BETA-D-GLUCANS
CC ALSO CONTAINING 1,3-LINKAGES.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY H (FAMILY 12 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D00546; BAA00435.1; -
DR EMBL: X52525; CAA36757.1; -
DR PIR: J00458; J00458.
DR PIR: S12610; S12610.
DR PIR: S14118; S14118.
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR ProDom: PD004316; Glyco_hydro.12; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; signal.
KW CELLULOSE DEGRADATION; HYDROLASE; POTENTIAL.
FT SIGNAL 1 16
FT CHAIN 17 237 ENDUGLUCANASE I.
FT MOD_RES 17 17 PYRROLIDONE CARBOXYLIC ACID.
SO SEQUENCE 237 AA; 25560 MW; 8f173571a8a6931 CRC64;

```

```

Query Match 14.7%; Score 213; DB 1; Length 237;
Best Local Similarity 31.1%; Pred. No. 2,9e-11;
Matches 73; Conservative 40; Mismatches 92; Indels 30; Gaps 12;

OY 40 ELGRMDARDVAGGRYVINNVWCAET--AACIEV--GLETFNFTTRDHONG--NNVA 93
Db 20 QLDQY--ATYTGVIYINNINMGKDGSGOCTYVNASASACTSNSTKNNWSGENSVK 77
OY 94 AYPAITFGCHWGACTSNSGL--PRVQELSDVTSWTLTPITTG--RMNAAVDIMFSPVT 149
Db 78 SY-----ANSGLTFENKLVISQIPTARMSYDNTGIRADVADLFTAAID 124
OY 150 NSNGYSGAELMIMNMGVMPGSSRAVTEVLGATWEVW-ADMNNYIAYRRTPT 208
Db 125 NHYT-WSGYELMTLARRGGVOPIGSOIATATVDQTELMWANGASOKYVSFAVPTPI 183
OY 209 TSVSELDLNAFIDDAVA-RGYIRPEWYLAHVETGFEIMEG-AGLSAPESVYQ 261
Db 184 TSF-QGDVNDVFYKLNQNHGFASQYLITLQPGTEPFGPATLSVSNWSASVQ 237

RESULT 3
GUN_ASPAC
ID GUN_ASPAC STANDARD: PRT: 239 AA.
AC Q12679;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A)
DE (Cellulase A) (Carboxymethylcellulase) (CMCase-I).
GN CEKA.
OS Aspergillus awamori (var. kawachi).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=105351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 4308;
RX MEDLINE=96059347; PubMed=7586029;
RA Sakamoto S., Tanura G., Ito K., Ishikawa T., Iwano K., Nishiya N.;
RT "Cloning and sequencing of cellulase cDNA from Aspergillus kawachi
RL and its expression in Saccharomyces cerevisiae.";
RL Curr. Genet. 27:435-439(1995).
CC -1- FUNCTION: HAS CARBOXYMETHYLCCELLULASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY H (FAMILY 12 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D12901; BAA02297.1; -
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR ProDom: PD004316; Glyco_hydro.12; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; signal.
KW CELLULOSE DEGRADATION; HYDROLASE; POTENTIAL.
FT SIGNAL 1 16
FT CHAIN 17 239 ENDUGLUCANASE A.
SO SEQUENCE 239 AA; 25769 MW; 16b0304a1a18276D0 CRC64;

```

```

Query Match 11.3%; Score 163.5; DB 1; Length 239;
Best Local Similarity 27.3%; Pred. No. 4,9e-07;
Matches 63; Conservative 38; Mismatches 97; Indels 33; Gaps 11;

OY 41 LGRMDARDVAGGRYVINNVWCAET--AACIEV--LETGNTTTRDHONG--NNVA 94
Db 41 LGRMDARDVAGGRYVINNVWCAET--AACIEV--LETGNTTTRDHONG--NNVA 94

```

```

Db 19 MCSGYDS-ASSPPYSVNQNLMEGYOCTGSCQYVVDKLSSGASWHTKWTMSGCEGVKS 76
QY 95 YPALYFGCHWGACTSNSGL---PRRVOELSDVRTSWTLPTITTG-RNNAAYDIWFSPVTN 150
Db 77 Y-----SNSGLTFPDKLVSDVSSIFTSVTMSODDTVQADVSDL-FTANA 122
QY 151 SGNGYSGAGELMTLNNNGVGMPCGSRVATVELAGATWVWY-----ADMDMNTIARRT 205
Db 123 DHATSSGDYELMTLARYGVOPYIGKQIAATYVGKSMEWYGTSTQAGAEOKTYSFVAG 182
QY 206 TPTTSVELDKAFIDDAVA-RGYIREPMYLAHETGFELEMGAGLRSAD 255
Db 183 SPINMSG-DIKDFPNLTIONQGFPASSOHLITLQCGTEPTGTGPAFTYD 232

```

```

RESULT 4
BGLAL_LACSK STANDARD; PRT; 625 AA.
AC 048846;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-galactosidase large subunit (EC 3.2.1.23) (lactase).
GN LACL.
OS Lactobacillus sakei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20017;
RX MEDLINE=96118231; PubMed=8574399;
RA Obst M., Meding E.R., Vogel R.F., Hammes W.P.;
RT "Two genes encoding the beta-galactosidase of Lactobacillus sake.";
RL Microbiology 141:3059-3066(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X82287; CA57730.1; -
DR HSSP: P00722; 1BGL.
DR InterPro: IPR001649; Glyco_hydro_2.
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
DR Pfam: PF02837; Glyco_hydro_2_N; 1.
DR PRINTS: PR00132; GLHYDRLASE2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW Hydrolase; Glycosidase.
FT ACT_SITE 465 465 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 533 533 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 533 533 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 625 AA; 72457 MW; C658A3C46136B886 CRC64;

```

Query Match 6.4%; Score 93; DB 1; Length 625;
 Best Local Similarity 17.9%; Pred. No. 1.5;
 Matches 44; Conservative 39; Mismatches 87; Indels 76; Gaps 11;

```

QY 53 GRVYINNV-----WGAEFAOCIEVGLFNGFTTRADHDNNGNNAAP---AIFGCHWG 105
Db 342 KRLVINGVNNHEHNPGRITTAEDAMADACQMRNNINAVRTSHVPRDLRFYNGC--- 398
QY 106 ACTNSGCLPRRVOLSDVRTSWTLPTITGRMN--AAVDIFSPVTNSG---NGYSGA 159

```

```

Db 399 ----DQAGITYMAETNLESRSQKMGAVEPSWNPVGSYDEMEATLDRATNFETPKNHV 455
QY 160 ELMIWLMNNGVMPGSGRVATVELAGATWVWYADMDMNTIARRTP----- 208
Db 456 SILFW-----SLGNESYAGSVLEKMA-----YKQODPTLVHYEGVERA 496
QY 209 ----TSVSEL-----DLKAFIDDAVARGYIREPMYLAHETG-----F 242
Db 497 PEKARTISDVESRYATAPAIKAYLADNAPOKPTICE-YMHDMGNSLGMQSYIDLLSOY 555
QY 243 ELMEGG 248
Db 556 DMVGG 561

```

```

RESULT 5
YL19_ARCFU STANDARD; PRT; 741 AA.
AC 028161;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2119 precursor.
GN AF2119.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleen H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Goceyne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arliach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AE000958; AAB89151.1; -
DR TIGR: AF2119; -
DR Hypothetical protein; Signal; Complete proteome.
KW SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 741 HYPOTHETICAL PROTEIN AF2119.
FT CHAIN 23 741 HYPOTHETICAL PROTEIN AF2119.
SQ SEQUENCE 741 AA; 82084 MW; E8C7543552231583 CRC64;

```

Query Match 6.4%; Score 93; DB 1; Length 741;
 Best Local Similarity 24.4%; Pred. No. 1.8;
 Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

```

QY 75 ETGNFTTRADHDNNGNNAAP-----TYFCHGACACTSNSG-----LPRR- 116
Db 204 DTDITLVYARVNSGPNVPPVYFAEFPDSQIFLID--GAWTNSGIGKLSFIPNVG 260
QY 117 -----VOELSDVTSWTLPTITGRNNAAYDIWFSP-----VTSNGN-GYSGAELM 162

```



```

FT CHAIN 17 1449 E2 GLYCOPROTEIN.
FT DOMAIN 17 1390 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1391 1410 POTENTIAL.
FT DOMAIN 1411 1449 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1411 1432 CVS-RICH.
FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 819 819 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1324 1324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1358 1358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1449 AA; 159957 MW; 971BBAE191FDIAF CRC64;

```

```

Query Match 6.0%; Score 87; DB 1; Length 1449;
Best Local Similarity 23.0%; Pred. No. 12;
Matches 45; Conservative 20; Mismatches 69; Indels 62; Gaps 10;

```

```

QY 45 WDARVAGGRVYNNVGAETACIEVLEGTGNTTRADHNGNNV-AAVPAIFYG-- 101
DB 88 WD-----YATENSTWNHK--QRLNVVNGVYPSIVTTRRNNSAGATICKGSP 137
QY 102 -----CHMGA-----CTNSGLPRVQELSDVRTSWTLPTTGRMNAAY 141
DB 138 PTTTSSLTCNMWSECLNHNKFPICPSNS-----EANGCMMLY 176
QY 142 DI-WESPVTNSGNGSGAELIMWL--NMNGVMPGSGRVATVELAGATWETWYWD--WD 196
DB 177 GLQWFAADAVVA---YLHGSYRISFENQWSGTVLLGDMKATTLTAGTLVLDLWMPYVD 233
QY 197 WNYIAYRRTPTTSVS 212
DB 234 VSYRVNKNKGTIVYS 249

```

```

RESULT 10
ABFB_STRCO STANDARD; PRT; 475 AA.
AC 054161;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
GN ABFB OR SC7H1.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D., Parkhill J., Barrell B.G., Raftandream M.A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC arabinofuranoside residues in alpha-L-arabinosides.
CC -1- PATHWAY: XYLAN DEGRADATION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 62 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
CC EMBL: AL021411; CA16189.1;
CC InterPro: IPR00772; Ricin_B_lectin.
CC Pfam: PF00652; Ricin_B_lectin; 1.
CC SMART: SM00458; RICIN; 1.
CC PROSITE: PS50231; RICIN_B_LECTIN; 1.
CC K1 xylan degradation; Hydrolase; Glycosidase; Signal; Lactin.
FT SIGNAL 1 37
FT DOMAIN 38 475 ALPHA-L-ARABINOFURANOSIDASE.
FT CHAIN 39 166 RICIN B-TYPE LECTIN.
SQ SEQUENCE 475 AA; 50045 MW; 47E707EE543CA60D CRC64;

```

```

Query Match 5.9%; Score 85; DB 1; Length 475;
Best Local Similarity 21.8%; Pred. No. 5.2;
Matches 53; Conservative 31; Mismatches 93; Indels 66; Gaps 14;

```

```

QY 11 SLLLFQGC-----DWLPFD-----GDN-----GKEPEPEPEPELQCGWRDARDVAG 52
DB 63 ALLQIDYDCWGTNQOWTSTDTGRITVYGDKCLDVGHATAPGTRVOI-----WSCSGGAN 117
QY 53 GRYRVYNN--VWGAETACIE-VGLEGTGNTTRADHNG-----NNVAA 94
DB 118 QQWRVNSDGTIVGVESGLCELAAGTAGTAVQLMTGCGGNGQWTLGTTGPTPDGCA 177
QY 95 YPAITFGCHMGACTISNGLPRVQELSDVRTSW-TLPITTGGRMNAAYDIWESPVTNSGN 153
DB 178 LPSTY---RW-----SSTGV-----LAQPKSGWVALKDFTTYTHNGRLVYGS--TSSGS 222
QY 154 GYSGGAELIMLNMNGVMPGSGRVATVELAGA-----TWYVYADMDMNNYIAYRRT 205
DB 223 SY-GSMVPSPTNMSDMSAGQANNAQAAVAPTLTFYFAPKNIVLWLYXQWGSMPFIYRTSS 281
QY 206 TPT 208
DB 282 DPT 284

```

```

RESULT 11
CHIT_STRPL
ID CHIT_STRPL STANDARD; PRT; 610 AA.
AC P11220;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Chitinase 63 precursor (EC 3.2.1.14).
GN CHTA.
OS Streptomyces plicatus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1922;
RN [1]

```



```

FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 832 832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1198 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1292 1292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1309 1309 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1334 1334 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1339 1339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1356 1356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1369 1369 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1447 AA; 160115 MW; 062EBD052DE2637 CRC64;

```

Query Match 5.8%; Score 84; DB 1; Length 1447;
 Best Local Similarity 23.4%; Pred. No. 22;
 Matches 51; Conservative 24; Mismatches 73; Indels 70; Gaps 12;

```

QY 37 PYVE---LCGRDADYAGGRVINV---WGAET-----AACIEVLETGNETT 82
DB 60 PIVQPFNCRINDSNDL---YVLENLKALYWDATENTNMHRLNVVNGYRST 115
QY 83 RADHDGNVVAAYPAIRFG-----CHMGA-----CTNSGLPRVOE 119
DB 116 VYTTTFNENAEGLIICICGSPPTTSSSLTCNMGSECLRNKRPICPSN----- 167
QY 120 LSDVTSWTLRPTTGRNMAAYDI-WFSPVTSNGNGSGAELMIL--NMNGVMPGGS 176
DB 168 -----EANGCMNLGLQWFA---DEVAYILHGASRYRSFENMGSGYTFPGDM 211
QY 177 RYATVELAGATWEVYAD--WDMNTIAYRRTPTTSVS 212
DB 212 RATTLVAGTLVDLWMFNPYDYVSYRVNKNKGTIVS 249

```

RESULT 13
 YBEM_ECOLI
 ID YBEM_ECOLI STANDARD; PRT; 468 AA.
 AC P75733;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ybfm.
 GN YBEM OR B0681.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino M., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";

```

RL DNA Res. 3:137-155(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000172; AAC73775.1;
CC EMBL; D90707; BAA35329.1; ALT_INIT.
CC Ecogene; EG13659; ybfm.
CC Hypothetical protein: Complete proteome.
SQ SEQUENCE 468 AA; 52780 MW; 954B3A778A61C2E4 CRC64;

```

Query Match 5.7%; Score 83; DB 1; Length 468;
 Best Local Similarity 23.6%; Pred. No. 7.5;
 Matches 46; Conservative 25; Mismatches 70; Indels 54; Gaps 10;

```

QY 83 RADHDGNVVAAYPAIRFGCHMGACTSNGGLPRVOELSDVTSWT----- 128
DB 281 RSYNDLYDGTAWLQALTFG-----YRADVDRLRLEGTWAKDGGQGYFLQR 327
QY 129 LPTTGRNMAAYDIWFSPVTSNGNGSGAELMIL-----NMNGVMPGGRVATV 181
DB 328 MTP-TYASSNGRLDLMW---DNRSDFNANGKAVFFGAYDLKWN---LPFGAIGASY 379
QY 182 ELAGATWEVYADWDMNTIAY--RRTPTTSVSELDLKAFLDADAVAGYIREPWLHAVE 239
DB 380 VYA---WDKAPATWQSNPDYADKNRTIESAVSLDAVYTTDGNAKGM---FHLHFE 433
QY 240 ----TGFELMEGAG 250
DB 434 YDHSIDIPSMGGYG 448

```

RESULT 14
 ABFB_STRLI
 ID ABFB_STRLI STANDARD; PRT; 475 AA.
 AC P96463;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
 GN ABFB.
 OS Streptomyces lividans.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1916;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-66 / 1326;
 RX MEDLINE=97220396; PubMed=9148759;
 RA Vincent P., Shareck F., Dupont C., Morosoli R., Kluepfel D.;
 RT "New alpha-L-arabinofuranosidase produced by Streptomyces lividans:
 RT cloning and DNA sequence of the abfb gene and characterization of the
 RT enzyme.";
 RL Biochem. J. 322:845-852(1997).
 RN [2]
 RP REVISIONS.
 RC STRAIN-1326;
 RA Shareck F.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: HAS A SPECIFIC ARABINOFURANOSE-DEBRANCHING ACTIVITY ON
 CC XYLAN FROM GRAMINAE. ACTS SYNERGISTICALLY WITH THE XYLANASES AND
 CC BINDS SPECIFICALLY TO XYLAN. FROM SMALL ARABINOXYLO-OLIGOSIDES,
 CC IT LIBERATES ARABINOSE AND, AFTER PROLONGED INCUBATION, THE
 CC PURIFIED ENZYME EXHIBITS SOME XYLANOXYLYTIC ACTIVITY AS WELL.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
 CC arabinofuranoside residues in alpha-L-arabinosides.
 CC -1- PATHWAY: XYLAN DEGRADATION.

CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO FAMILY 62 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M64551; AAC26524.1; -
 DR InterPro: IPR000772; Ricin_B_lectin.
 DR Pfam: PF00652; Ricin_B_lectin; 1.
 DR SMART: SM00458; RICIN; 1.
 DR PROSITE: PS50231; RICIN_B_LECTIN; 1.
 RM Xylan degradation; Hydrolase; Glycosidase; Signal; Lactin.
 FT SIGNAL 1 37
 FT CHAIN 38 475 ALPHA-L-ARABINOFURANOSIDASE.
 FT DOMAIN 39 166 RICIN B-TYPE LECTIN.
 SO SEQUENCE 475 AA; 50369 MW; C3CB14EE7BF85AD CRC64;

Query Match 5.7%; Score 83; DB 1; Length 475;
 Best Local Similarity 22.4%; Pred. No. 7.7;
 Matches 55; Conservative 31; Mismatches 88; Indels 72; Gaps 15;

OY 11 SLLEFGC-----DMLPPD-----GDN-----GKEPEPEPTVELCGRWADRVAG 52
 DB 63 ALLQLIDCWGNTQOWTSTDTGRLTVGDKLDVPGHATAPETRQIWSG-----SG 114
 OY 53 GR---YRVINN--VWGAEATACIE-VGLETFNFTTRADHDNG-----NN 91
 DB 115 GRNQOQVRVNSDGTGVGVSGCLTEAAGAGTPNGTAVOLMTGCGGNQKWTGLTGPPTDG 174
 OY 92 VAATPAITFGCHMGACTNSGLPRVQELSDVRTSM-TLPTTTRKMAADYIMSPVTN 150
 DB 175 TCALPSTY---RW---SSTGV---LAOPKSGVALKDETTVTHNGRLHYGS--TS 219
 OY 151 SGNGYSGAEIMIMNMNGVMPGSRVATVELAGA-----TWYVYADMDNNTAY 202
 DB 220 SSSST--GSMVTSPTFNMSDMASAGQMANQAAVATLTFYFAKNIWLAAYONGSPFIYR 278
 OY 203 RRTPT 208
 DB 279 TSSDPT 284

RESULT 15
 GUXB_CELFI STANDARD; PRT; 1090 AA.
 AC P50899;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellulohydrolase B)
 DE (1,4-beta-cellulohydrolase B) (CBP120).
 GN CBHB OR CENE.
 OS Cellulomonas fimi.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
 OX NCB1_taxonomy:1708;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
 RC STRAIN-ATCC 484;
 RX MEDLINE-96003898; PubMed-7575482;
 RA Shen H., Giles N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RT "Cellulohydrolase B, a second exo-cellulohydrolase from the
 RT cellulytic bacterium Cellulomonas fimi."
 RL Biochem. J. 311:67-74(1995).
 RN [2]
 RP SEQUENCE OF 54-75.

RX MEDLINE-93209933; PubMed-8458833;
 RA Meinke A., Giles N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 RT D (Cend), a family A beta-1,4-glucanase.";
 RL J. Bacteriol. 175:1910-1918(1993).
 RN [3]
 RP SEQUENCE OF 54-78.
 RX MEDLINE-94197708; PubMed-8147863;
 RA Shen H., Yomme P., Meinke A., Giles N.R., Kilburn D.G.,
 RA Warren R.A.J., Miller R.C. Jr.;
 RT "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi
 RT Cend, a member of a new family of beta-1,4-glucanases.";
 RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
 CC -1- FUNCTION: HYDROLYSE CELLULOSE TO A MIXTURE OF CELLOTRIOSE,
 CC CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT
 CC HYDROLYSED CELLOPENTAPOSE TO CELLOTRIOSE AND CELLOBIOSE, AND
 CC CELLOTRIOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.
 CC HAS ALSO WEAK ENDOLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS
 CC WITH INVERSION OF ANOMERIC CONFIGURATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and celloetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 CC (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L38827; AAB00822.1; -
 DR HSSP: P07986; IEXG.
 DR InterPro: IPR001919; CBD_2.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003962; FN_III_repeat.
 DR InterPro: IPR000556; Glyco_hydro_48.
 DR Pfam: PF00553; CBD_2; 1.
 DR Pfam: PF00041; fn3; 3.
 DR Pfam: PF02011; Glyco_hydro_48; 1.
 DR PRINTS: PR00014; FNTYPEIII.
 DR PRINTS: PR00844; GLHYDRASE48.
 DR PRODOM: PD011903; Glyco_hydro_48; 1.
 DR SMART: SM00060; FN3; 3.
 DR PROSITE: PS00561; CBD_BACTERIAL; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 33
 FT PROPEP 34 53
 FT CHAIN 54 1090
 FT DOMAIN 54 699 EXOGLUCANASE B.
 FT DOMAIN 700 785 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 794 884 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 891 978 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 989 1090 FIBRONECTIN TYPE-III 3.
 FT ACT_SITE 513 513 CELLULOSE-BINDING (BY SIMILARITY).
 FT DISULFID 990 1089 NUCLEOPHILE (BY SIMILARITY).
 SO SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;

Query Match 5.7%; Score 82.5; DB 1; Length 1090;
 Best Local Similarity 23.8%; Pred. No. 21;
 Matches 41; Conservative 12; Mismatches 55; Indels 67; Gaps 8;

OY 45 WDARDVAGR-----YRVINWGAETACIEVLEGTN-FTTRADHDNGNNVAAP 96
 DB 915 WNASDTFGSGLKGYDYRGATRVGTTTASVTDGLRAATAYQYTVATDNAGVSA-- 972
 OY 97 AITFGCHMGACTNSGLPRVQELSDVRTSM-TLPTTTRKMAADY--WFS-----PV 148

Db 973 -----ASAALSVTKPTPOTGGSCSVAYNASSMNSGFTASVRI 1009
Oy 149 TNSG---NGYSGAEL-----MTWLN-----WNGGVMPG 174
Db 1010 TNGTTTNGWSLGFDLTAGOKVOQGSATWTOSGSTVTATNAPWNCTLAPG 1061

Search completed: August 15, 2002, 10:58:02
Job time: 380 sec


```
Db 61 VMAETACIEVEGLETGNFTTRADHDNGNNVAIPATITGCHMAPRAIRIDCAARGAV 120
QY 115 RVOELSDVTSWTLPTITGRNNAAYDIWSPVTNSGNGSGAEELIMLNNGGVP 174
Db 121 RRAHELD-----VPEITGRNNAAYDIWSPVTNSGNGSGAEELIMLNNGGVP 173
QY 175 GSRVAVVELAGATWVWYADMDNVIYRRTPTTSVELDKAFIDAVARGIRPEWY 234
Db 174 GSRVAVVELAGATWVWYADMDNVIYRRTPTTSVELDKAFIDAVARGIRPEWY 233
QY 235 LHAETGFELEMGAGLRSDAFSTV 261
Db 234 LHAETGFELEMGAGLRSDAFSTV 260

RESULT 2
ID 09KIH1 PRELIMINARY; PRT; 371 AA.
AC 09KIH1:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELULASE 12A.
GN CEL12A.
OS Streptomyces sp. 11A68.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_Taxid=13452;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-11A68;
RA van Solingen P., Meijer D., van der Kleij W.A.H., Barnett C.C.,
RA Bolle R., Power S.D., Jones B.E.;
RT "Cloning and expression of an endocellulase gene from a novel
RT Streptomyces isolated from an East African soda lake.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233376; AAF91283.1; -
DR HSSP; P07986; 1EXG.
DR InterPro; IPR002594; Glyco_hydro_12.
DR InterPro; IPR001230; Prenyltn.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR PROSITE; PS00294; PRENYLATON; UNKNOWN_1.
SQ SEQUENCE 371 AA; 38481 MW; 0E1BC4288A148914 CRC64;

Query Match 27.4%; Score 396.5; DB 2; Length 371;
Best Local Similarity 39.7%; Pred. No. 6.7e-24;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 40 ELGGRDADVAGGRVYNNVGAETACIEVEGLETGN-FTTRADHDNGNNVA--AYP 96
Db 34 QICDRGTTTID-DRVYVONNRNGTSATQICV--TGNGFETIOADGSPRTGAPKSY 89
QY 97 AIFGCHMGACTNSGGLPRVOELSDVTSWTLPTITGRNNAAYDIWSPVTNSGNGS 156
Db 90 SYVDGCHGNCAPRTTLPRKISSIGSAPSSVSRITGNGVYNAADIDWDP--TPRTN 148
QY 157 GGAELMIMLNNGGVPVGGSRVAVVELAGATWVWYADMDNVIYRRTPTTSVELD 216
Db 149 -RTEIMIMNRVGPVOPISGPGTAHVGRSMEVWTGSGNSDVISFLAPSAISSM 207
QY 217 KAFIDAVARGIRPEWYLHAETGFELEMGAGLRSDAFSTV 260
Db 208 KQFVDVAVSHGLATPDWLTSLQAGFEPEWEGGTGLAVNSSFSAV 251

RESULT 3
ID 059963 PRELIMINARY; PRT; 382 AA.
AC 059963:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
```

```
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELULASE (EC 3.2.1.4).
GN EGLS.
OS Streptomyces rochei (Streptomyces parvulus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_Taxid=1928;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-A2;
RA MEDLINE=95011642; PubMed=7523249;
RA Perito B., Hanhart E., Irdani T., Iqbal M., McCarthy A.J.,
RA Mastromel G.;
RT "Characterization and sequence analysis of a Streptomyces rochei A2.";
RL Gene 148:119-124(1994).
DR EMBL; X73953; CA52139.1; -
DR HSSP; P07986; 1EXG.
DR InterPro; IPR001919; CBD_2.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF00553; CBD_2; 1.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Hydrolyase; Glycosidase.
SQ SEQUENCE 382 AA; 39398 MW; 21C014342EFC6565 CRC64;

Query Match 24.3%; Score 352; DB 2; Length 382;
Best Local Similarity 36.2%; Pred. No. 2.5e-20;
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

QY 50 VAGGRVYNNVGAETACIEVEGLETGNFTTRADHDNGNNVA--AYPAIFGCHMGAC 107
Db 49 VIGGRVYVONNRNGTSATQCV-TATSG-FRVTQAGSVPTGAPKSYSVNGCHYTNC 106
QY 108 TNSGGLPRVOELSDVTSWTLPTITGRNNAAYDIWSPVTNSGNGSGAEELIMLN 167
Db 107 SPETALPARISGISAPSSISYGFVONAVNASYDIDLPPTPTDGS--VNRTIEMIFNR 164
QY 168 NGVMPGSRVAVVELAGATWVWYADMDNVIYRRTPTTSVELDKAFIDAVARG 227
Db 165 VGOIQIGSQVGTASVAGRTWEVSGGKNTDLSVAPASMSNMFVMDVRAVANG 224
QY 228 YIRPEWYLHAETGFELEMGAGLRSDAFSTV 260
Db 225 LAGNDWLTSLQAGFEPEWONGAGLAVNSSFSTV 257

RESULT 4
ID 008468 PRELIMINARY; PRT; 377 AA.
AC 008468:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CEL2 (EC 3.2.1.4).
GN CELA2.
OS Streptomyces halstedii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_Taxid=1944;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN-JM8;
RA MEDLINE=97307849; PubMed=9182697;
RA Garda-Salas A.L., Fernandez-Abalos J.M., Sanchez P., Ruiz-Arribas A.,
RA Santamaria-Sanchez R.I.;
RT "Two genes encoding an endoglucanase and a cellulose-binding protein
RT are clustered and co-regulated by a TTA codon in Streptomyces
RT halstedii JM8.";
RL Biochem. J. 324:403-411(1997).
DR EMBL; U51222; AAC45429.1; -
```



```

QY 171 VMBGSRVATVELAGATWEVYADWDMNYTAYRRTPPTTSVSELDLKAFTDDAVARCYIR 230
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 171 IQPIGSPVGNASVGGRTWMEVSGNGSDNYLSFVAPSAISGSPFVDMDFRATVARGLAE 230
QY 231 PEWYIHAVETGFELMEGGAGLRSDAFSVTYO 261
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 231 NDKYLTISVQAGFEPMQNGAGLAVNSFSSTVE 261

RESULT 7
Q9X602 PRELIMINARY; PRT; 384 AA.
ID O9X602;
AC O9X602;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CELLULASE.
GN CEL51.
OS Streptomyces viridosporus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=67581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=77A;
RA Ramachandran S., Crawford D.L.;
RT "Characterization and sequence analysis of two genes involved in
RT cellulase degradation in Streptomyces viridosporus 77A, and its
RT expression in Escherichia coli.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF130408; AND25090.1; -.
DR HSSP: P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
SQ SEQUENCE 384 AA; 40918 MW; D3968B6EBDEDE65 CRC64;

Query Match 23.4%; Score 338.5; DB 2; Length 384;
Best Local Similarity 32.1%; Pred. No. 3e-19;
Matches 85; Conservative 35; Mismatches 106; Indels 39; Gaps 8;

QY 31 PEPEP-----EPTVELGGRDARDVA-----GGRYRYNNWGAETA 67
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 7 PDPAFCVAVSPSPSPSSRRRRAQADPTLCEPYGTTTIOGRYVYQNNRMGSSSP 66
QY 68 QCIEVGLFTGNFTITRADHDNGNNA--AYPATYFGCHGACSTNSGLPRVQELSDVRT 125
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 67 QCV-TATDTG-FRLTQADGSPVTPNGAPKSTPSVFNCHYTNCSPGTRKLPARISGISAP 124
QY 126 -----SWTLPTITGR----WNAAYDIFSPYTSNGSGGAEIMIMLNNGVMPG 175
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 125 RIXGISASSPSSISYGVGAVYNAAYDILMDPTPTDG--VNTEIMINENKVKPIQPIG 182
QY 176 SRVATVELGATWEVYADWDMNYTAYRRTPPTTSVSELDLKAFTDDAVARCYIRPEWY 235
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 176 SRVATVELGATWEVYADWDMNYTAYRRTPPTTSVSELDLKAFTDDAVARCYIRPEWY 235
QY 183 SOVGTAIVTGGRFQWVWVGSGNGSDNYLSFVAPSAIESWSEPDVDFRETVARGMQNDWYL 242
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 183 SOVGTAIVTGGRFQWVWVGSGNGSDNYLSFVAPSAIESWSEPDVDFRETVARGMQNDWYL 242
QY 236 HAVETGFELMEGGAGLRSDAFSVTY 260
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 236 HAVETGFELMEGGAGLRSDAFSVTY 260
QY 243 TSVQAGFEPMQNGAGLAVNSFSSTV 267
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 243 TSVQAGFEPMQNGAGLAVNSFSSTV 267

RESULT 8
Q31030 PRELIMINARY; PRT; 264 AA.
ID O31030;
AC O31030;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

```

```

DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
GN CELB.
OS Pectobacterium carotovorum subsp. carotovorum.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LY34;
RX MEDLINE=98096373; PubMed=9434760;
RA Park Y.W., Lim S.T., Cho S.J., Yun H.D.;
RT "Characterization of Erwinia carotovora subsp. carotovora 1734 endo-
RT 1,4-beta-glucanase genes and rapid identification of their gene
RT products.";
RL Biochem. Biophys. Res. Commun. 241:636-641(1997).
DR EMBL: AF025769; AAC02965.1; -.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
KW Signal; Hydrolase.
FT SIGNAL 1 36 POTENTIAL.
FT CHAIN 37 264 BETA(1,4)-GLUCAN GLUCANOHYDROLASE.
SQ SEQUENCE 264 AA; 29634 MW; 70EA366B8443CEB CRC64;

Query Match 16.8%; Score 242.5; DB 2; Length 264;
Best Local Similarity 29.8%; Pred. No. 8.6e-12;
Matches 68; Conservative 39; Mismatches 90; Indels 31; Gaps 9;

QY 54 RRYVNNWGAETVACQIEVGLFTGNFTITRAD-----HDGNNVAAATPATYFGCHGKA 106
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 48 KYVFNNWVGKDEVK---GMOQTFVYNSPTSGMWHMPPSSSVKAYPSLVSQWMTA 103
QY 107 C-TSNGSLPRVQELSDVRTSWTLPTITGRWNAAYDIFSPYTSNGSGGAEIMIML 165
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 104 YTEENSGLPKILKSNKSIINSVYTSIKSTATLYAAADVHFHTDKASMSIPDELMIML 163
QY 166 NMNGVMPGSGSRVATVELAGATWEVYADW-----DMNYTAYRRTPPTTSVSELDLKA 218
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 164 N-NTNAGPAGDYETVFLDSSNNV-FKGMINAGNGGNNVFSVTSNTNSAS-LNRH 220
QY 219 FIDDAVARCYIRPEW-----YLHAVETGFELMEGGAGLRSDAFSVTYO 261
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 221 FTMDVQ-----KKWMSNAKYISSVELGTEIRFGDGQIDITKMSDYK 264

RESULT 9
Q9KYS4 PRELIMINARY; PRT; 244 AA.
ID O9KYS4;
AC O9KYS4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE SECRETED SUGAR HYDROLASE.
GN SC5H4.15.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;

```


RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinasht H., Hopwood D.A.:
 RA "A set of ordered cosmids and a detailed genetic and physical map for
 RA the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL355913; CAB91127.1.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase.
 SQ SEQUENCE 244 AA; 26123 MW; 2CA63242798HEAD4 CRC64;

Query Match 14.1%; Score 204.5; DB 2; Length 244;
 Best Local Similarity 26.7%; Pred. No. 8.5e-09;
 Matches 60; Conservative 52; Mismatches 78; Indels 35; Gaps 11;

OY 42 CGRMARVAGRGYRVINNVWGAET-AOCI--EVLGTGNFTITRADHDGNNVAAPAI 98
 DB 46 CDQWNTSLNG--YTLVNNWGSAGSGOCVANSSTDMGVW---ADHPNTDGISYSP-- 97
 OY 99 YFGCHMGACTSNGLPRVOELSDVTSWTLTPITGRNNAADIMFSPVTSNGMGYSGG 158
 DB 98 -----NAKKVINKPITSLSSLTSSVNTVPSSGAANTSYDIDM-----DTDDY---- 140
 OY 159 AELMVLWNGVMPGSGRVATVELAGATWEVYVADMDNNTI-AVRRTPPTS--VSELD 215
 DB 141 -EIMLMVNHGAVGVLGFGQSGVGLGHTWVDYKGNNGANEFSEFLRTSDSNSTVNIILP 199
 OY 216 LKAFIDDAVARGYIRPEWYLHAVETGTFELMEGAGGLRSADFSVTY 260
 DB 200 ILKWLKD--TKGMGNE-TIGDVOFGYEITSSAGGLDFTNNLTV 241

RESULT 10
 ID 060033 PRELIMINARY: PRT: 274 AA.
 AC 060033.
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULOSE)
 DE (ENDOGLUCANASE) (CARBOXYMETHYL CELLULOSE).
 GN CELB OR TM1525.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 (DSM3109);
 RX MEDLINE=96425879; PubMed=8828221;
 RA Liebl W., Rulle P., Bronnenmeier K., Riedel K., Lotzsch F.,
 RA Greif I.:
 RA "Analysis of a Thermotoga maritima DNA fragment encoding two similar
 RT thermostable cellulases, CelA and CelB, and characterization of the
 RT recombinant enzymes.";
 RL Microbiology 142:2532-2542(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwynn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Lihner K.D., Garrett M.M.,
 RA Steward A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.:
 RA "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 DR EMBL: Z69341; CAA93274.1; -.
 DR EMBL: AE001800; AAD36592.1; -.
 DR TIGR: TM1525; -.

DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 KW Signal; Hydrolase; Glycosidase; Complete proteome.
 FT SIGNAL 1 17
 FT CHAIN 18 274
 SQ SEQUENCE 274 AA; 31734 MW; 72C0A5ACAD7BA358 CRC64;

Query Match 14.0%; Score 202; DB 16; Length 274;
 Best Local Similarity 28.3%; Pred. No. 1.5e-08;
 Matches 73; Conservative 40; Mismatches 87; Indels 58; Gaps 16;

OY 38 TVELGRMDARDVAGRGYRVINNVWGAETACIEVLEGTNFTITRADHDGNNVAAPAI 97
 DB 38 TVEL-NFNVKSYEG-----ETWLKFDGEKVEFYADLYNNLONPD-----SVHGYPE 85
 OY 98 IYFGCH-WGACTSNNG--LPRVOELSD--VRTSWTL-----PTITGRNNAADIMF 146
 DB 86 IYGYKPMAG--HNSGVEFLPVKRDLPDYVYLDYSLWYENNLPI-----NLAEWTIT 138
 OY 147 PVTNNGNGYSGGAEELMVLWNGVMPGSGRV-----ATVELAG-----ATWEVYADMDN 198
 DB 139 RSPDQTSVSSGDALIMWF-YNNVLMPPGQKVDFTTYEINGVKQETKMDYFAPMGMD 197
 OY 199 YTAIRRTPTTSVS-ELDIKAFIDDA---VARGYIRPEWYLHAVETGTFELMEGA----- 249
 DB 198 YLAFRLTTPMKEGKAKINVKDFQRAAEVKKHSTRID---NFEELYCYWBEIGTEFGDP 254
 OY 250 -----GLRSADFSVTY 260
 DB 255 NTKAKFGWTFRDSEYEV 272

RESULT 11
 ID P96492 PRELIMINARY: PRT: 274 AA.
 AC P96492.
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ENDO-1,4-BETA-GLUCANASE B (EC 3.2.1.4) (CELLULOSE) (ENDOGLUCANASE)
 DE (CARBOXYMETHYL CELLULOSE).
 GN CELB.
 OS Thermotoga neapolitana.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC BOK J.D., Eveleigh D.E.:
 RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-137 FROM N.A.
 RC STRAIN=Z2706-MC24;
 RA Zverlov V.:
 RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-137 FROM N.A.
 RC STRAIN=Z2706-MC24;
 RX MEDLINE=93356813; PubMed=8352795;
 RA Dakhova O., Kurepina N., Zverlov V., Svetlichny I V.,
 RA Velikovskaya G.:
 RA "Cloning and expression in Escherichia coli of Thermotoga neapolitana
 RT genes coding for enzymes of carbohydrate substrate degradation.";
 RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
 CC -I- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 LINKAGES IN CELLULOSE.
 DR EMBL: U93354; AAC95060.1; -.
 DR EMBL: Z86103; CAB06782.1; -.
 DR InterPro: IPR002594; Glyco_hydro_12.
 DR Pfam: PF01670; Glyco_hydro_12; 1.
 DR ProDom: PD004316; Glyco_hydro_12; 1.
 KW Hydrolase; Glycosidase.

01-DEC-2001 (TREMELREL. 19, last annotation update)
DE ENDO-BETA-1,4-GLUCANASE (EC 3.2.1.4).
GN EGL.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM9414;
RA Okada H., Tada K., Sekiya T., Yokoyama K., Takahashi A., Tonda H.,
Kumagai H., Morikawa Y.,
RT Molecular characterization and heterologous expression of the gene
RT encoding a low-molecular-mass endoglucanase from Trichoderma reesei
RT QM9414.
RL Appl. Environ. Microbiol. 64:55-563 (1998).
DR EMBL: AB003694; BAA20140.1;
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR Prodom: PD004316; Glyco_hydro_12; 1.
KW Hydrolyase; Glycosidase.
SQ SEQUENCE 234 AA; 25159 MW; DF476EDE384ADD1 CRC64;

Query Match 12.2%; Score 176.5; DB 3; Length 234;
Best Local Similarity 28.6%; Pred. No. 1.4e-06;
Matches 68; Conservative 34; Mismatches 93; Indels 43; Gaps 14;

OY 42 CGRMDARDVAGRGYRVINNWGAETQ---CI-EVGLGTGNTITRAD---HDNGNNVAA 94
DB 20 CDQM--ATTGNGITYVSNLMDGASAGSGGCTVAVSLSG--ASWHADQWQSGQNNVKS 75
OY 95 YPAIVEGCHWGAETNSGLP--RRVQELSDVPT--SWTLPTTGRNNAAYDIEMFPTN 150
DB 76 YQ-----NSQIAIPQKRTVNSISSMPTTASWSYSG--SNIRANVAYDL-FTANP 122
OY 151 SGNGYSGGAEIMTLMWNGGVMPGSRVATVELAGATWEVWY----ADWDNTIATIRRT 206
DB 123 NHTYSGDDELMTLKGYPGIGIPISGQGVNNGQSWTLGYNGAMQVYSFVAQTNIT 182
OY 207 PPTSVELDLKAFI---DDAVARGYIRPEWYLHAVETGFELMEGAGLRSADPSVT 260
DB 183 NYSG----DYKNFPTNLRDN---KGTNAAGQVLYSTQFTGTEPTGSGTLNVAASVTASI 233

RESULT 15

ID 013454 PRELIMINARY; PRT; 239 AA.
AC 013454;
DT 01-JAN-1998 (TREMELREL. 05, Created)
DT 01-JAN-1998 (TREMELREL. 05, last sequence update)
DT 01-DEC-2001 (TREMELREL. 19, last annotation update)
DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).
GN CELA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KBN616;
RX MEDLINE=97161783; PubMed=9008887;
RA Kitamoto N., Go M., Shibayama T., Kimura T., Kito Y., Ohmura K.,
Tsukagoshi N.,
RT Molecular cloning, purification and characterization of two endo-1,4-
RT beta-glucanases from Aspergillus oryzae KBN616.
RL Appl. Microbiol. Biotechnol. 46:538-544 (1996).
DR EMBL: D83731; BAA22588.1;
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR Prodom: PD004316; Glyco_hydro_12; 1.
KW Hydrolyase; Glycosidase.
SQ SEQUENCE 239 AA; 26096 MW; C0F850E5DFEB455D CRC64;

Query Match 11.9%; Score 172.5; DB 3; Length 239;
Best Local Similarity 27.3%; Pred. No. 2.9e-06;
Matches 63; Conservative 37; Mismatches 86; Indels 45; Gaps 11;
OY 40 ELGCRMDARDVAGRGYRVINNWGAET---AOCIEVGLGTGNTITRADHDNGNNVAA 96
DB 18 ELCAQYDS--ASSPPYSVNNLMDGQSGTSCQCVY-----DNLSSGA--59
OY 97 AITFGCHW-----GACTSNSGLPRRYQEL--SDVRTSWTLPTTGRNNAAYDI 143
DB 60 AMHTTWTWNGEGSVKSYNSAVTFDKLIVDSQSIPTDVMQDNTNNA---DVAAYDL 116
OY 144 WFSPTNSGNGYSGAEELMTLMWNGGVMPGSRVATVELAGATWEVWY----ADWDNN 198
DB 117 -FTAADQNNHTYISGDYELMTLMARYSTIQIDTATAYEGHTWELWGTGTTQAGAEQK 175
OY 199 YIAYRRTPTTSVELDLKAFIDDAVAR-GYIRPEWYLHAVETGFELMEG 248
DB 176 TYSFVSATPINTFGG-DIKKFDYITSKHSFPASAQYLLINMGTEPPTFG 225

Search completed: August 15, 2002, 10:57:39
Job time: 393 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:51:03 ; Search time 53.28 Seconds

(without alignments)
462.807 Million cell updates/sec

Title: us-10-003-759-2_COPY_40_261

Perfect score: 1234
Sequence: 1 ELGGRMDARDVAGGRYVIN.....FEIMEGAGLRSDFSVTYQ 222

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802:*

```
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1106.5	89.7	260	20	AAV06369 Rhodothermus marinus
2	1106.5	89.7	260	21	AA18482 Emmericella descripta
3	1106.5	89.7	260	21	AAV84347 Amino acid sequenc
4	470.5	38.1	105	20	AAV06342 Rhodothermus marinus
5	396.5	32.1	371	20	AAV06367 Streptomyces sp. 3
6	396.5	32.1	371	20	AAV08473 Actinomyces sp. E
7	396.5	32.1	371	21	AA18480 Streptomyces livid
8	396.5	32.1	371	21	AAV84345 Amino acid sequenc
9	396.5	32.1	371	21	AAV67496 Actinomyces cell
10	396.5	32.1	386	21	AAV67497 Protein sequence o
11	343	27.8	381	20	AAV06368 Streptomyces livid

12	343	27.8	381	21	AA18481 Rhodothermus marinus
13	329	26.7	429	21	AAV84346 Amino acid sequenc
14	255.5	20.7	261	20	AAW88462 Bacillus lichenifo
15	252.5	20.5	264	20	AAV06370 Erwinia carotovora
16	252.5	20.5	264	21	AA18483 Amino acid sequenc
17	252.5	20.5	264	21	AAV84348 Amino acid sequenc
18	224.5	18.2	259	20	AAV06351 Aspergillus aculea
19	224.5	18.2	259	21	AA18464 Amino acid sequenc
20	224.5	18.2	259	21	AAV84329 Aspergillus kawach
21	214	17.3	239	21	AAV06353 Aspergillus kawach
22	214	17.3	239	21	AA18466 Amino acid sequenc
23	214	17.3	239	21	AAV84331 Humicola insolens
24	212	17.2	234	21	AA18469 Humicola insolens
25	212	17.2	254	21	AAV84334 Amino acid sequenc
26	210.5	17.1	246	20	AAV06366 Emmericella descripta
27	210.5	17.1	246	21	AA18479 Actinomyces sp. E
28	210.5	17.1	246	21	AAV84344 Amino acid sequenc
29	209	16.9	254	21	AA18468 Humicola grisea EG
30	209	16.9	254	21	AAV84333 Humicola grisea EG
31	209	16.9	254	21	AAV07558 Humicola grisea en
32	209	16.9	254	22	AAU07584 Humicola grisea en
33	208.5	16.9	254	22	AAU07583 Humicola grisea en
34	201.5	16.3	255	20	AAV06356 Humicola insolens
35	199	16.1	104	20	AAV06337 Streptomyces sp. E
36	192.5	15.6	253	20	AAV06355 Humicola grisea EG
37	192.5	15.6	348	21	AAV06362 Gliocladium roseum
38	192.5	15.6	348	21	AA18475 Amino acid sequenc
39	192.5	15.5	234	22	AAV84340 Trichoderma reesei
40	191.5	15.5	234	22	AAU07582 Tiarospora phas
41	184.5	15.0	244	19	AAW68593 T. phaseolina xylo
42	184.5	15.0	244	21	AAV44341 Fusarium javanicum
43	183	14.8	244	20	AAV06359 Fusarium javanicum
44	183	14.8	244	21	AA18472 Amino acid sequenc
45	183	14.8	244	21	AAV84337

ALIGNMENTS

RESULT 1	AAV06369	AAV06369 standard; Protein; 260 AA.
ID	AAV06369	
XX	AAV06369	
AC	AAV06369	
DT	06-SEP-1999 (first entry)	
XX		
DE	Rhodothermus marinus EGII-1-like cellulase.	
XX		
KW	Cellulase; endoglucanase; EGII-1; textile; feed additive; baking; food processing; grain wet milling; pulp; paper.	
XX		
OS	Rhodothermus marinus.	
XX		
PN	WO9931255-A2.	
XX		
PD	24-JUN-1999.	
XX		
PF	14-DEC-1998; 98WO-US26552.	
XX		
PR	16-DEC-1997; 97US-0991720.	
XX		
PA	(GENV) GENENCOR INT INC.	
XX		
PI	Bower BS, Fowler T, Phillips JT;	
XX		
DR	WPI, 1999-395187/33.	
XX		
PT	EGII-1 like cellulase	
XX		
PS	Example; Fig 6; 47pp; English.	
XX		
CC	The present polypeptide represents a full-length sequence of a	

CC novel EgitII-like cellulase of Rhodothermus marinus. It was
CC deduced from a gene sequence isolated from genomic DNA using PCR
CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)
CC of Trichoderma reesei EgitII cellulase and related enzymes. PCR
CC has been used to identify novel EgitII-like enzymes, including the
CC present protein, from bacterial and fungal sources (see AAY06331-70).
CC The sequence shows homology to T. reesei EgitII (see AAY06330). Also
CC provided by the invention are vectors, host cells and methods
CC for the recombinant production of such enzymes, which can be used
CC in the treatment of cellulose-containing textiles, as feed
CC additives, in the treatment of wood pulp, in the reduction of
CC biomass to glucose, in the stone washing of indigo dyed denim, or
CC as laundry detergent components (all claimed).

XX Sequence 260 AA:

XX The present sequence is a cellulase related to endoglucanase III (EgitII)
XX from Trichoderma reesei. EgitII-like genes were isolated from genomic DNA
XX libraries constructed from various microorganisms by PCR. The isolated
XX genes showed significant homology to EgitII from T. reesei. Certain
XX substitution and deletion mutations have been incorporated into EgitII and
XX EgitII-like cellulases to produce variant enzymes with improved stability,
XX e.g. increased resistance to temperature stress. The mutants may be used
XX in textile and wood pulp treatment, as a feed additive, and for reducing
XX biomass to glucose. They are also useful for stonewashing or indigo dyed
XX denim and as an agent in laundry and dish detergents.

XX Sequence 260 AA:

Query Match 89.7%; Score 1106.5; DB 20; Length 260;
Best Local Similarity 88.6%; Pred. No. 9.6e-107;
Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

OY 1 ELGGRDARDVAGRRYVNNVGAETACIEVGLTGNFTTRADHDGNNVAAVPAIY 60

OY 1 ELGGRDARDVAGRRYVNNVGAETACIEVGLTGNFTTRADHDGNNVAAVPAIY 60

Db 40 elggrdardvagrgrvynnwgaetaclegvletgnftltradhngnnvaapaly 99

Db 40 elggrdardvagrgrvynnwgaetaclegvletgnftltradhngnnvaapaly 99

OY 61 FCGHMKR-----CTNSGLPRVQELSDVRSWTLPITTTGRNNAAYDIWSPVTSNGN 114

OY 61 FCGHMKR-----CTNSGLPRVQELSDVRSWTLPITTTGRNNAAYDIWSPVTSNGN 114

Db 100 fgcwparairdcaaragavrraheld-----vplltgrwnaaydlwfsptvnsn 152

Db 100 fgcwparairdcaaragavrraheld-----vplltgrwnaaydlwfsptvnsn 152

OY 115 GYGGAELMIMLNMGVMPGSRVATVELAGATWEVYADMDMNYIAYRTPTTSVSE 174

OY 115 GYGGAELMIMLNMGVMPGSRVATVELAGATWEVYADMDMNYIAYRTPTTSVSE 174

Db 153 gysggaelmimlwnvgvmpgsrvatvelagatwewyadwknjlayrrtpttsvse 212

Db 153 gysggaelmimlwnvgvmpgsrvatvelagatwewyadwknjlayrrtpttsvse 212

OY 175 LDKAFIDDAVANGYIRPEWYLHAVETGFEWEGAGLSADSFVTVQ 222

OY 175 LDKAFIDDAVANGYIRPEWYLHAVETGFEWEGAGLSADSFVTVQ 222

Db 213 ldlkafiddavargylrpewylhavetgfelwegagltradsfvtrvq 260

Db 213 ldlkafiddavargylrpewylhavetgfelwegagltradsfvtrvq 260

RESULT 2

RESULT 3

ID AAB14882 standard; Protein: 260 AA.

ID AAY84347 standard; Protein: 260 AA.

AC AAB14882;

AC AAY84347;

DT 21-NOV-2000 (first entry)

DT 12-JUL-2000 (first entry)

DE Emeritella desertoru EgitII-like cellulase.

DE Amino acid sequence of an endoglucanase III (EgitII)-like cellulase.

KW Emeritella desertoru; Trichoderma reesei; endoglucanase III; EgitII;
KW cellulase; mutant; enzyme stability; textile treatment;

KW Endoglucanase III; EgitII; EgitII-like cellulase; surfactant stability;
KW cellulase; textile processing; textile cleaning; stonewashing;
KW indigo dyed denim; cellulose containing fabric; fabric smoothness;
KW p111 removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;
KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.

OS Emeritella desertoru.

OS Rhodothermus marinus.

PN WO200037614-A2.

PN WO200014208-A1.

PD 29-JUN-2000.

PD 16-MAR-2000.

PF 12-NOV-1999; 99WO-US26704.

PF 24-AUG-1999; 99WO-US19154.

PR 18-DEC-1998; 98US-0216295.

PR 03-SEP-1998; 98US-0146729.

PA (GENEV) GENENCOR INT INC.

PA (GENEV) GENENCOR INT INC.

PI Mitchinson C, Wendt DJ;

PI Fowler T;

DR WPI: 2000-482483/42.

DR WPI: 2000-271052/23.

PT Novel endoglucanase III or endoglucanase III-like cellulase useful for
PT treating textiles and wood pulp comprises a substitution or deletion at
PT specified positions in the wild form of endoglucanase III -
XX
XX
XX Example 1; Fig 3; 52pp; English.

PT Novel variant endoglucanase III-like cellulases with improved
PT surfactant stability and resistance to temperature stress, useful for
PT textile processing or cleaning, treating wood pulp, food and grain, and
PT reducing biomass to glucose -
XX
XX

PS Disclosure; Page 66-67; 73pp; English.
 XX
 CC The present sequence represents an endoglucanase III (EgIII)-like
 CC cellulase. The cellulase has homology to the Trichoderma reesei EgIII
 CC protein. The variant cellulases have improved temperature stability,
 CC and improved surfactant stability. The variant cellulases and
 CC compositions containing them are used in textile processing or cleaning,
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
 CC or appearance of cellulose containing fabrics (e.g. improving fabric
 CC smoothness or removing pills and fibrils). The compositions may also be
 CC used for the removal of immature or dead cotton from cellulosic fibres
 CC or fabric, which can cause uneven drying. The cellulase may also be used
 CC in a detergent composition for washing laundry and dishes and in the
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
 CC The enzymes may also be used in the reduction of biomass to glucose.
 CC
 XX
 SO Sequence 260 AA:

Query Match 89.7%; Score 1106.5; DB 21; Length 260;
 Best Local Similarity 88.6%; Pred. No. 9.6e-107;
 Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;
 QY 1 ELGGRMDARVAGGRYVNNWGAETACIEVGLTGNFTTRADHDGNNVAAPAIY 60
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
 40 elggrwardvaggrrvnnwgaetaqcievgletgnftltradhngnnvaapaily 99
 QY 61 FCGHMG-----CTNSGLPRRVOELSDVFTSWLTPITGRRNNAAYDIWFSPTNSGN 114
 Db ||||| : : ||||| : ||||||||||||||||||||||||
 100 fgcHWaparalrdcaaragavrraheld-----vtpiltgrnaaydiwfsptnsgn 152
 QY 115 GYSGAGELMTLNNNGVMEGSGRATVELAGATWEWYADMDNNTAYRRTTTSYSE 174
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
 153 gysggaelmwlnwngvmpgsrvalagatwewyadmdwnyayrrtpttsvse 212
 QY 175 LDKAFIDDAVARGYRPEWTLHAVETGFELEMGAGLRSDFSVTVO 222
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
 213 ldkafidavargyrirpeytlhavetgfelwegaglrtsdsvtqv 260

RESULT 4
 AAY06342
 ID AAY06342 standard; Protein: 105 AA.
 XX
 AC AAY06342;
 XX
 DT 06-SEP-1999 (first entry)
 XX
 DE Rhodothermus marinus EgIII-like cellulase (partial sequence).
 XX
 KW Cellulase; endoglucanase; EgIII; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 XX
 OS Rhodothermus marinus.
 XX
 PN W09931255-A2.
 XX
 PD 24-JUN-1999.
 XX
 PF 14-DEC-1998; 98WO-US26552.
 XX
 PR 16-DEC-1997; 97US-0991720.
 XX
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Bower BS, Fowler T, Phillips JT;
 XX
 DR WPI; 1999-395187/33.
 XX
 PT EgIII like cellulase
 XX
 PS Example; Fig 3; 47pp; English.
 XX

CC The present polypeptide represents a partial sequence of a novel
 CC EgIII-like cellulase of Rhodothermus marinus. It was deduced from
 CC a partial gene sequence isolated from genomic DNA using PCR
 CC primers (see AX59180-91) based on conserved motifs (see AAY06325-29)
 CC of Trichoderma reesei EgIII cellulase and related enzymes. PCR
 CC has been used to identify novel EgIII-like enzymes, including the
 CC present polypeptide, from bacterial and fungal sources (see
 CC AAY06331-70). Also provided by the invention are vectors, host
 CC cells and methods for the recombinant production of such enzymes,
 CC which can be used in the treatment of cellulose-containing textiles,
 CC as feed additives, in the treatment of wood pulp, in the reduction
 CC of biomass to glucose, in the stone washing of indigo dyed denim,
 CC or as laundry detergent components (all claimed).
 CC
 XX
 SO Sequence 105 AA:

Query Match 38.1%; Score 470.5; DB 20; Length 105;
 Best Local Similarity 77.7%; Pred. No. 3.6e-41;
 Matches 87; Conservative 3; Mismatches 9; Indels 13; Gaps 2;
 QY 20 NNWGAETACIEVGLTGNFTTRADHDGNNVAAPAIYFCGHMG-----CTNSG 73
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
 1 nnwgaetaqcievgletgnftltradhngnnvaapailyfgcHWaparalrdcaarag 60
 QY 74 LPRRVOELSDVFTSWLTPITGRRNNAAYDIWFSPTNSGNGSGAELMTW 125
 Db ||||| : : ||||| : ||||||||||||||||||||||||
 61 avrraheld-----vtpiltgrnaaydiwfsptnsgngysggaelmw 105

RESULT 5
 AAY06367
 ID AAY06367 standard; Protein: 371 AA.
 XX
 AC AAY06367;
 XX
 DT 06-SEP-1999 (first entry)
 XX
 DE Streptomyces sp. EgIII-like cellulase.
 XX
 KW Cellulase; endoglucanase; EgIII; textile; feed additive; baking;
 KW food processing; grain wet milling; pulp; paper.
 XX
 OS Streptomyces sp.
 XX
 PN W09931255-A2.
 XX
 PD 24-JUN-1999.
 XX
 PF 14-DEC-1998; 98WO-US26552.
 XX
 PR 16-DEC-1997; 97US-0991720.
 XX
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Bower BS, Fowler T, Phillips JT;
 XX
 DR WPI; 1999-395187/33.
 XX
 PT EgIII like cellulase
 XX
 PS Example; Fig 6; 47pp; English.
 XX
 CC The present polypeptide represents a full-length sequence of a
 CC novel EgIII-like cellulase of Streptomyces sp. 11A68. It was
 CC deduced from a gene sequence isolated from genomic DNA using PCR
 CC primers (see AX59180-91) based on conserved motifs (see AAY06325-29)
 CC of Trichoderma reesei EgIII cellulase and related enzymes. PCR
 CC has been used to identify novel EgIII-like enzymes, including the
 CC present protein, from bacterial and fungal sources (see AAY06331-70).
 CC The sequence shows homology to T. reesei EgIII (see AAY06330). Also
 CC provided by the invention are vectors, host cells and methods
 CC for the recombinant production of such enzymes, which can be used

biomass to glucose. They are also useful for stonewashing or indigo dyed denim and as an agent in laundry and dish detergents.

Sequence 371 AA;

Query Match 32.1%; Score 396.5; DB 21; Length 371;
Best Local Similarity 39.7%; Pred. No. 1e-32;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGMARDVAGGRYRVNNVWGAETACIEVGLGTGN-FTTRADHDNGNNA--AYP 57
Db 34 qtdcrgygtlttlq-dryvvqnmrwgtsatqclnv---tgnfcltqadgsvprtnqpkxyp 89
QY 58 AIYFCGHCAGTSNSGLPRROELSDVRTSWTLPTITGRNMAAYDIFSPYTNNGNGYS 117
Db 90 svydgchygncaprtllpmrlssigsapsvsyrytgnvyaaaydiwldp-tptrngvn 148
QY 118 GGAELMIMLNMGVMPGSGRYATVELAGATWEVYADMDMNYIARRTPPTTSVELDL 177
Db 149 -rteimlwnrnyvqpvqpgyspvglahvggrswewtcsgnsgndvisflapsaiswscfv 207
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELMEGAGLRSADEFSYTV 221
Db 208 kdfvdqavshglatpdytltsiqagfepwegtclavnsfssav 251

RESULT 8
AAV84345
ID AAV84345 standard; Protein; 371 AA.
AC AAV84345;

DT 12-JUL-2000 (first entry)

DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.

XX Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;
KM cellulase; textile processing; textile cleaning; stonewashing;
KM indigo dyed denim; cellulose containing fabric; fabric smoothness;
KM pill removal; fibril removal; cotton; cellulosic fibre; dyeing; detergent;
KM animal feed; wood pulp; paper; grain; biomass reduction; glucose.

XX Actinomyces sp.

OS WO200014208-A1.

PN 16-MAR-2000.

XX 24-AUG-1999; 99WO-US19154.

XX 03-SEP-1998; 98US-0146729.

XX (GENEV) GENENCOR INT INC.

XX Fowler T;

XX WPI: 2000-271052/23.

PT Novel variant endoglucanase III-like cellulases with improved
PT surfactant stability and resistance to temperature stress, useful for
PT textile processing or cleaning, treating wood pulp, food and grain, and
PT reducing biomass to glucose

PS Disclosure; Page 64-65; 73pp; English.

XX The present sequence represents an endoglucanase III (EGIII)-like
CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII
CC protein. The variant cellulases have improved temperature stability,
CC and improved surfactant stability. The variant cellulases and
CC compositions containing them are used in textile processing or cleaning,
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC or appearance of cellulose containing fabrics (e.g. improving fabric

CC smoothness or removing pills and fibrils). The compositions may also be
CC used for the removal of immature or dead cotton from cellulosic fibres
CC or fabric, which can cause uneven dyeing. The cellulase may also be used
CC in a detergent composition for washing laundry and dishes and in the
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
CC The enzymes may also be used in the reduction of biomass to glucose.

Sequence 371 AA;

Query Match 32.1%; Score 396.5; DB 21; Length 371;
Best Local Similarity 39.7%; Pred. No. 1e-32;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGMARDVAGGRYRVNNVWGAETACIEVGLGTGN-FTTRADHDNGNNA--AYP 57
Db 34 qtdcrgygtlttlq-dryvvqnmrwgtsatqclnv---tgnfcltqadgsvprtnqpkxyp 89
QY 58 AIYFCGHCAGTSNSGLPRROELSDVRTSWTLPTITGRNMAAYDIFSPYTNNGNGYS 117
Db 90 svydgchygncaprtllpmrlssigsapsvsyrytgnvyaaaydiwldp-tptrngvn 148
QY 118 GGAELMIMLNMGVMPGSGRYATVELAGATWEVYADMDMNYIARRTPPTTSVELDL 177
Db 149 -rteimlwnrnyvqpvqpgyspvglahvggrswewtcsgnsgndvisflapsaiswscfv 207
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELMEGAGLRSADEFSYTV 221
Db 208 kdfvdqavshglatpdytltsiqagfepwegtclavnsfssav 251

RESULT 9
AAV67496
ID AAV67496 standard; Protein; 371 AA.
AC AAV67496;

DT 19-MAY-2000 (first entry)

DE Actinomyces cellulase protein sequence.

XX Cellulase; Actinomyces; detergent; feed additive; textile treatment;
KM pulp; paper.

XX Streptomyces sp.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Key Location/Qualifiers
FT Peptide 1..27
FT Protein /note= "signal sequence"
FT /note= "mature protein"
PN WO200009707-A1.
XX 24-FEB-2000.
PD 28-MAY-1999; 99WO-US11971.
PF 24-JUN-1998; 98US-0104308.
PR 18-NOV-1998; 98WO-US24649.
PR 28-MAY-1999; 98US-0321981.
XX (GENEV) GENENCOR INT INC.
XX Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
XX WPI: 2000-224344/19.
DR N-PSDB; AA257029.

XX A novel Actinomyces cellulase and related DNA, useful for detergent
PT compositions, treating textiles and paper or pulp
XX Claim 1; Fig 1; 72pp; English.

CC cellulase. The cellulase has homology to the *Trichoderma reesei* Egit
CC protein. The variant cellulases have improved temperature stability,
CC and improved surfactant stability. The variant cellulases and
CC compositions containing them are used in textile processing or cleaning,
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel
CC or appearance of cellulose containing fabrics (e.g. improving fabric
CC smoothness or removing pills and fibrils). The compositions may also be
CC used for the removal of immature or dead cotton from cellulosic fibres
CC or fabric, which can cause uneven dyeing. The cellulase may also be used
CC in a detergent composition for washing laundry and dishes and in the
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.
CC The enzymes may also be used in the reduction of biomass to glucose.
XX
XX
SQ Sequence 429 AA:

Query Match 26.7%; Score 329; DB 21; Length 429;
Best Local Similarity 35.0%; Pred. No. 1.3e-25;
Matches 75; Conservative 32; Mismatches 101; Indels 6; Gaps 4;

QY 11 VAGGRVYNNVGAETACIEVLEGTGNTTTRADHNGNNA--AYPAIYGCWAGAC 68
DB 100 vfgchytqnmrvgsatpqcvtatdtg-ftrvqadgsaptngakpsvfnghytnct 157
QY 69 TSNSGLPRVQELSDVRTSWTLPTTGRWNAAYDIMFSPVTSNGSGAGFLMIMLN 128
DB 158 spgtdlprvldtvsapssisfyvgdavyasaydlwloptardg--vntqetlmwfnr 215
QY 129 NGGVMPGSGRVATVELAGATWEVYADMDWNYIAYRRTPTTSVSELDKAFIDDAVARQ 188
DB 216 vspidqldgspvgasvgrtewvsgngsndvlsfvpalsgwsfdvmdfrratvarg 275
QY 189 YIREPWYLAHVEFGFELMEGAGLRSDPSVTYQ 222
DB 276 laendwyltsvqagfepwngaglavnsfsstve 309

RESULT 14
AAW88462
ID AAW88462 standard; Protein; 261 AA.
XX
XX AAW88462;
XX
XX 10-MAY-1999 (first entry)
XX
XX DE Bacillus licheniformis xyloglucanase.
XX
XX KW xyloglucanase; detergent.
XX
XX OS Bacillus licheniformis.
XX
XX FH Key Location/Qualifiers
XX FT Protein 30..261
XX FT /note= "mature protein, this region is specifically
XX claimed in Claim 19"

W09902663-A1.
XX
XX 21-JAN-1999.
XX
XX PD 01-JUL-1998; 98WO-DK00290.
XX
XX PF 24-OCT-1997; 97DK-0001213.
XX PR 07-JUL-1997; 97DK-0000822.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX PA Bjornvad ME, Jorgensen PL, Outtrup H, Schuelein M;
XX PI WPI: 1999-120866/10.
XX DR N-PSDB; AAX06949.
XX
XX PT New enzyme preparation comprising a xyloglucanase with an activity

PT of 50 % at pH 7 - useful for improving the properties of cellulosic
PT fibres, yarn, (non)woven fabric, and rating hemp, jute, flax and
PT linen fibres
XX
XX Claim 29; Page 71-72; 87pp; English.

CC This polypeptide is an alkaline xyloglucanase obtained from
CC Bacillus licheniformis ATCC 14580. The enzyme shows optimal
CC activity at 60 deg C, and retains 50% of its activity after 20 min
CC at 70 deg. The optimum pH for activity is 5.54, with 32% relative
CC activity retained at pH 4.59, and 45% at pH 8.49. The heat is
CC 16.5/sec on xyloglucan at pH 7.5, Km 1.1 g/l. The ratio of maximum
CC xyloglucanase activity to maximum activity on CM-cellulose is at
CC least 5:1. An isolated polynucleotide (see AAX06949) encoding the
CC xyloglucanase can be utilised in the production of recombinant
CC enzyme. Xyloglucanase preparations are useful for improving the
CC properties of cellulosic fibres, yarn, (non)woven fabric, and for
CC rating hemp, jute, flax and linen fibres (claimed). They can also
CC be used in a process for machine treatment of fabrics, and in
CC detergent compositions (claimed). The enzyme preparations exhibit
CC high xyloglucanase activity at alkaline pH without essentially
CC attacking cellulose or cellulose derivatives.
XX
XX
SQ Sequence 261 AA:

Query Match 20.7%; Score 255.5; DB 20; Length 261;
Best Local Similarity 30.6%; Pred. No. 2.8e-18;
Matches 70; Conservative 41; Mismatches 85; Indels 33; Gaps 9;

QY 15 RYVYNNVWGAETACIEVLEGTGNTTTRADHD-----NKNNAAYPAIYF 61
DB 45 kyllfnnvsgadqvgwqgli-----yhnsdsdmgwwnpsntstvkaypsivs 94
QY 62 GCHM-GACTSNGLPRVQELSDVRTSWTLPTTGRWNAAYDIMFSPVTSNGSGAGA 120
DB 95 gwhwtegytsgsfprlisdqknlnkvsysangtynaaydlvlnlnkswsdaptcd 154
QY 121 ELMWLNMGVWPGGSRVATVELAGATWEV--YAD---MDWNYIAYRRTPTTSVSE 174
DB 155 elmlwin-ntnagpagsyvetvsgshwkykygyladagsgkymvnsftrancgs-en 212
QY 175 LDKAFID-AYARGYIREPWYLAHVEFGFELMEGAGLRSDPSVTYQ 222
DB 213 lnrdfnyladsqkwlstkkyssvefglevfygldqinsnwdvtvr 261

RESULT 15
AAV06370
ID AAV06370 standard; Protein; 264 AA.
XX
XX AAV06370;
XX
XX 06-SEP-1999 (first entry)
XX
XX DE Erwania carotovora Egit-like cellulase.
XX
XX KW Cellulase; endoglucanase; Egit; textile; feed additive; baking;
XX food processing; grain wet milling; pulp; paper.
XX
XX OS Erwania carotovora.
XX
XX PN W09931255-A2.
XX PD 24-JUN-1999.
XX
XX PF 14-DEC-1998; 98WO-US26552.
XX PR 16-DEC-1997; 97US-0991720.
XX
XX (GENEV) GENENCOR INT INC.
XX
XX PA Bower BS, Fowler T, Phillips JT;
XX
XX PT

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:52:08 ; Search time 23.32 Seconds
(without alignments)
232.525 Million cell updates/sec

Title: US-10-003-759-2_COPY_40_261
1234
Sequence: 1 ELGCRMDARDVAGGRYRVIN.....FELMEGAGLRSDPSVTYQ 222

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1106.5	89.7	260	4	US-09-216-295-23 Sequence 23, Appl
2	396.5	32.1	312	4	US-09-216-295-21 Sequence 21, Appl
3	396.5	32.1	371	4	US-09-104-308-1 Sequence 1, Appl
4	396.5	32.1	371	4	US-09-321-981-1 Sequence 1, Appl
5	396.5	32.1	386	4	US-09-321-981-5 Sequence 5, Appl
6	343.27	27.8	381	4	US-09-216-295-22 Sequence 22, Appl
7	255.5	20.7	261	4	US-09-110-959A-2 Sequence 2, Appl
8	254	20.6	233	1	US-08-032-848C-12 Sequence 12, Appl
9	246	19.9	263	4	US-09-216-295-24 Sequence 24, Appl
10	224.5	18.2	259	4	US-09-216-295-5 Sequence 5, Appl
11	214	17.3	239	4	US-09-216-295-7 Sequence 7, Appl
12	213	17.3	221	1	US-08-032-848C-13 Sequence 13, Appl
13	212	17.2	253	4	US-09-216-295-10 Sequence 10, Appl
14	210.5	17.1	246	4	US-09-216-295-20 Sequence 20, Appl
15	209	16.9	253	4	US-09-216-295-9 Sequence 9, Appl
16	192.5	15.6	348	4	US-09-216-295-16 Sequence 16, Appl
17	183	14.8	243	4	US-09-216-295-13 Sequence 13, Appl
18	179.5	14.5	232	4	US-09-146-770-1 Sequence 1, Appl
19	176.5	14.3	218	1	US-08-032-848C-10 Sequence 10, Appl
20	176.5	14.3	218	1	US-08-438-870-10 Sequence 10, Appl
21	176.5	14.3	218	2	US-08-169-948B-34 Sequence 34, Appl
22	176.5	14.3	218	2	US-08-448-873-34 Sequence 34, Appl
23	176.5	14.3	218	4	US-08-382-452D-34 Sequence 34, Appl
24	176.5	14.3	218	4	US-09-216-295-1 Sequence 1, Appl
25	176.5	14.3	234	1	US-08-032-848C-9 Sequence 9, Appl
26	176.5	14.3	234	1	US-08-438-870-9 Sequence 9, Appl
27	176.5	14.3	234	4	US-09-146-770-3 Sequence 3, Appl

28	176.5	14.3	234	4	US-09-216-295-3	Sequence 3, Appl
29	172	13.9	239	4	US-09-216-295-15	Sequence 15, Appl
30	167	13.5	237	4	US-09-216-295-19	Sequence 19, Appl
31	166.5	13.5	234	4	US-09-146-770-4	Sequence 4, Appl
32	166.5	13.5	234	4	US-09-216-295-4	Sequence 4, Appl
33	163.5	13.2	239	4	US-09-216-295-6	Sequence 6, Appl
34	163.5	13.2	250	4	US-09-216-295-14	Sequence 14, Appl
35	163	13.2	247	4	US-09-216-295-8	Sequence 8, Appl
36	162.5	13.2	223	3	US-08-913-264-1	Sequence 1, Appl
37	162.5	13.2	239	3	US-08-913-264-1	Sequence 13, Appl
38	161.5	13.1	239	4	US-08-849-751-2	Sequence 2, Appl
39	161.5	13.1	239	4	US-09-478-816-2	Sequence 2, Appl
40	160.5	13.0	319	3	US-09-215-042-1	Sequence 1, Appl
41	159.5	12.9	288	4	US-09-216-295-12	Sequence 12, Appl
42	140	11.3	288	3	US-08-995-280C-2	Sequence 2, Appl
43	140	11.3	288	4	US-09-215-042-2	Sequence 2, Appl
44	138	11.2	194	4	US-09-216-295-11	Sequence 11, Appl
45	106.5	8.6	368	3	US-08-591-685-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-216-295-23
; Sequence 23, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EGII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 23
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Rhodothermus marinus
; US-09-216-295-23

Query Match 89.7%; Score 1106.5; DB 4; Length 260;
Best local Similarity 88.6%; Pred. No. 3e+102;
Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;
QY 1 ELGCRMDARDVAGGRYRVINNWGAETACIEVLEGTGNTTRADHDGNNVAAYPATY 60
DB 40 ELGCRMDARDVAGGRYRVINNWGAETACIEVLEGTGNTTRADHDGNNVAAYPATY 99
QY 61 FGCHMGA-----CTNSGILPRRVOELSDVRTSWLTLPITTGNNAAVDIWFSPVNSGN 114
DB 100 FGCHMGA-----CTNSGILPRRVOELSDVRTSWLTLPITTGNNAAVDIWFSPVNSGN 152
QY 115 GYGGAGELMTLWLNWNGVMPGSGRVATVELAGATWEVYADMDMNTIARTPTTSVSE 174
DB 153 GYGGAGELMTLWLNWNGVMPGSGRVATVELAGATWEVYADMDMNTIARTPTTSVSE 212
QY 175 LDKAFIDDAVANGYRPEWYLAHAVETGPELMEGAGLRSDPSVTYQ 222
DB 213 LDKAFIDDAVANGYRPEWYLAHAVETGPELMEGAGLRSDPSVTYQ 260
RESULT 2
US-09-216-295-21
; Sequence 21, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328el Variant EGII-Like Cellulase Compositions
; FILE REFERENCE: GC555

;; CURRENT APPLICATION NUMBER: US/09/216,295
;; CURRENT FILING DATE: 1998-12-18
;; NUMBER OF SEQ ID NOS: 41
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 21
;; LENGTH: 312
;; TYPE: PR1
;; ORGANISM: Actinomyces 11AG8
US-09-216-295-21

Query Match 32.1%; Score 396.5; DB 4; Length 312;
Best Local Similarity 39.7%; Pred. No. 1.2e-31;

Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDARDVAGGRVYINNNWGAETACIEVEGLETGN-FTTRADHDGNVNA--AYP 57
DB 34 QICDRYGTITIQ-DRYVYONNRNGTSATOCINV---TGNGFELTQADGSPVPTNGAKRSYP 89
QY 58 AIYFGCHWGACTNSGLPRRVOELSDVRTSWTLPTTGRMNAAYDIWSPVTSNGNGYS 117
DB 90 SYVDGCHYGCAPRTLLPMMISSIGSAPSSVSRYTGNNGYNAAYDIWLDP-TPTNGVN 148
QY 118 GGAELMIMLNMGVAPGSGRVATVELAGATWEVWYADMDMYIARRPTTTSVELDL 177
DB 149 -RTEIMWENRNGVPVOPISPGTAHVGRSMWETGSGNSNDVISFLAPSAISSWSFDV 207
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELEMGAGLSADPSVTY 221
DB 208 KQVDQAVSHGLATPDWYLTISIAGFEPEWGGTGLAVNSFSAY 251

RESULT 3
US-09-104-308-1
Sequence 1, Application US/09104308
Patent No. 6187577

GENERAL INFORMATION:

APPLICANT: Jones, Brian E.
APPLICANT: Van Der Kleij, Wilhelmus A.H.

APPLICANT: Van Solingen, Piet

APPLICANT: Weyler, Walter

TITLE OF INVENTION: No. 6187577el Cellulase Producing Actinomyces,

TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International, Inc.

STREET: 925 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1013

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/104,308

FILING DATE: 24-JUN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/974,042

FILING DATE: 19-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Stone, Christopher L.

REGISTRATION NUMBER: 35,696

REFERENCE/DOCKET NUMBER: GC539

TELEPHONE: 650-846-7555

TELEFAX: 650-845-6504

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 371 amino acids

TYPE: amino acid

;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-09-104-308-1

Query Match 32.1%; Score 396.5; DB 4; Length 371;
Best Local Similarity 39.7%; Pred. No. 1.5e-31;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDARDVAGGRVYINNNWGAETACIEVEGLETGN-FTTRADHDGNVNA--AYP 57
DB 34 QICDRYGTITIQ-DRYVYONNRNGTSATOCINV---TGNGFELTQADGSPVPTNGAKRSYP 89
QY 58 AIYFGCHWGACTNSGLPRRVOELSDVRTSWTLPTTGRMNAAYDIWSPVTSNGNGYS 117
DB 90 SYVDGCHYGCAPRTLLPMMISSIGSAPSSVSRYTGNNGYNAAYDIWLDP-TPTNGVN 148
QY 118 GGAELMIMLNMGVAPGSGRVATVELAGATWEVWYADMDMYIARRPTTTSVELDL 177
DB 149 -RTEIMWENRNGVPVOPISPGTAHVGRSMWETGSGNSNDVISFLAPSAISSWSFDV 207
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELEMGAGLSADPSVTY 221
DB 208 KQVDQAVSHGLATPDWYLTISIAGFEPEWGGTGLAVNSFSAY 251

RESULT 4

US-09-321-981-1
Sequence 1, Application US/09321981

Patent No. 6287839

GENERAL INFORMATION:

APPLICANT: Genencor International, Inc.

TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomyces,

TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same

FILE REFERENCE: GC540-2

CURRENT APPLICATION NUMBER: US/09/321,981

CURRENT FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: PCT/US99/11971

PRIOR FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: 09/104,308

PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 08/974,042

PRIOR FILING DATE: 1997-11-19

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FASTSEQ for Windows Version 3.0

SEQ ID NO 1

LENGTH: 371

TYPE: PR1

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Nearest "neighbor" - Streptomyces

US-09-321-981-1

Query Match 32.1%; Score 396.5; DB 4; Length 371;
Best Local Similarity 39.7%; Pred. No. 1.5e-31;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDARDVAGGRVYINNNWGAETACIEVEGLETGN-FTTRADHDGNVNA--AYP 57
DB 34 QICDRYGTITIQ-DRYVYONNRNGTSATOCINV---TGNGFELTQADGSPVPTNGAKRSYP 89
QY 58 AIYFGCHWGACTNSGLPRRVOELSDVRTSWTLPTTGRMNAAYDIWSPVTSNGNGYS 117
DB 90 SYVDGCHYGCAPRTLLPMMISSIGSAPSSVSRYTGNNGYNAAYDIWLDP-TPTNGVN 148
QY 118 GGAELMIMLNMGVAPGSGRVATVELAGATWEVWYADMDMYIARRPTTTSVELDL 177
DB 149 -RTEIMWENRNGVPVOPISPGTAHVGRSMWETGSGNSNDVISFLAPSAISSWSFDV 207
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELEMGAGLSADPSVTY 221
DB 208 KQVDQAVSHGLATPDWYLTISIAGFEPEWGGTGLAVNSFSAY 251

Db 208 KDFVDAVSHGLATPDMYLTISQAGFEPMEGTCGLAVNSFSSAV 251

RESULT 5
US-09-321-981-5
Sequence 5, Application US/09321981

Patent No. 6287839

GENERAL INFORMATION:
APPLICANT: Genencor International, Inc.
TITLE OF INVENTION: No. 6287839e1 Cellulase Producing Actinomycetes,
CELLULOSE PRODUCED THEREFROM AND METHOD OF PRODUCING SAME
FILE REFERENCE: GC540-2

CURRENT APPLICATION NUMBER: US/09/321,981

CURRENT FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: PCT/US99/11971

PRIOR FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: 09/104,308

PRIOR FILING DATE: 1998-06-24

PRIOR FILING DATE: 1997-11-19

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 386

TYPE: PRF

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Nearest "neighbor" = Streptomyces

US-09-321-981-5

Query Match

Best Local Similarity 32.1%; Score 396.5; DB 4; Length 386;

Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

1 ELGGRMARVAGRGVYNNVWGAETACCEVLEGTGN-FTIRADHDNNVA--AYP 57

49 QICDRIGTTTQ-DRYVONNRKMTSATQCLNV--TGNGEITQADGSPVTPGAPKSY 104

58 AIYFGCHMGACTSNGLPRVVOELSDVTSWTLPTTGRMNAAYDIWFSPTNSGNGS 117

105 SYVDGCHYGCAPRTLLPMLISSIGSAPSSVRYTGNQYNNAYDIWLDPTRTGCVN 163

118 GGAELMTLWNGVMPGSRVATVELAGATWEVYADMDNNTAYRRTPTTSVSELD 177

164 -RREIMTFNRVGPVOPIGSPVGAHVGGRSMWEYWTGNSGNDVLSFLAPSAISSMSFDV 222

178 KAFIDDAVARGYRPEWYLAHVGTEFELMEGAGLRSADFSVTV 221

223 KDFVDAVSHGLATPDMYLTISQAGFEPMEGTCGLAVNSFSSAV 266

RESULT 6

US-09-216-295-22

Sequence 22, Application US/09216295

Patent No. 6268328

GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin

TITLE OF INVENTION: No. 6268328e1 Variant Egit-Like Cellulase Compositions

FILE REFERENCE: GC555

CURRENT APPLICATION NUMBER: US/09/216,295

CURRENT FILING DATE: 1998-12-18

NUMBER OF SEQ ID NOS: 41

SOFTWARE: FastSeq for Windows Version 3.0

Query Match

Best Local Similarity 27.8%; Score 343; DB 4; Length 381;

Matches 77; Conservative 31; Mismatches 97; Indels 6; Gaps 4;

14 GRYVINNWGAETACCEVLEGTGNFTIRADHDNNVA--AYPAIYFGCHMGACTSN 71

55 GRYVONNRKMTSATQCLNV--TGNGEITQADGSPVTPGAPKSY 104

72 SGLPVVOELSDVTSWTLPTTGRMNAAYDIWFSPTNSGNGSGAELMTLWNGG 131

113 TDLPVRLDYSAPSSISYGFVQAGVYNNAYDIWLDPTRTGCVN-VMQTEIMTFNRVGP 170

132 VMEGSRVATVELAGATWEVYADMDNNTAYRRTPTTSVSELDKAFIDDAVARGYR 191

171 IOPIGSPVGAHVGGRSMWEYWTGNSGNDVLSFLAPSAISSMSFDV 222

192 PEWYLAHVGTEFELMEGAGLRSADFSVTV 221

231 KDFVDAVSHGLATPDMYLTISQAGFEPMEGTCGLAVNSFSSAV 266

RESULT 7

US-09-110-959A-2

Sequence 2, Application US/09110959A

Patent No. 6268197

GENERAL INFORMATION:
APPLICANT: Schielein, Martin

APPLICANT: Outtrup, Helle

APPLICANT: Bjornvad, Mads Eskelund

TITLE OF INVENTION: Alkaline xylolucanase

FILE REFERENCE: 5206.200-US

CURRENT APPLICATION NUMBER: US/09/110,959A

CURRENT FILING DATE: 1998-07-07

PRIOR APPLICATION NUMBER: 0822/97

PRIOR FILING DATE: 1997-07-07

PRIOR APPLICATION NUMBER: 1213/97

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/054,039

PRIOR FILING DATE: 1997-07-28

PRIOR APPLICATION NUMBER: 60/063,694

PRIOR FILING DATE: 1997-10-28

NUMBER OF SEQ ID NOS: 14

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 261

TYPE: PRF

ORGANISM: Bacillus licheniformis ATCC 14580

US-09-110-959A-2

Query Match

Best Local Similarity 20.7%; Score 255.5; DB 4; Length 261;

Matches 70; Conservative 41; Mismatches 85; Indels 33; Gaps 9;

15 RYRIVNWGAETACCEVLEGTGNFTIRADHDNNVAAYPAIY 61

45 KYXIFNNWGAETACCEVLEGTGNFTIRADHDNNVAAYPAIY 94

62 GCHW-GACTSNGLPRVVOELSDVTSWTLPTTGRMNAAYDIWFSPTNSGNGSGA 120

95 GWHTEGTAGSGFPRLSDOKNINTKVSISANGRYNNAAYDIWLDPTRTGCVN-PTD 154

121 ELMTLWNGVMPGSRVATVELAGATWEVYADMDNNTAYRRTPTTSVSELD 174

```

US-08-032-848C-12
: Sequence 12, Application US/08032848C
: Patent No. 5475101
:
: GENERAL INFORMATION:
:   APPLICANT: Ward, Michael
:   APPLICANT: Clarkson, Kathleen A.
:   APPLICANT: Weiss, Geoffrey L.
:   APPLICANT: Larenas, Edward
:   APPLICANT: Lorch, Jeffrey D.
:   TITLE OF INVENTION: Purification and Molecular Cloning of
:   TITLE OF INVENTION: Eg III Cellulase
:   NUMBER OF SEQUENCES: 20
:   CORRESPONDENCE ADDRESS:
:   STREET: 180 Kimball Way
:   CITY: South San Francisco
:   STATE: CA
:   COUNTRY: USA
:   ZIP: 94080
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/032.848C
:   FILING DATE: MAR 17 1993
:   CLASSIFICATION: 435
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Horn, Margaret A.
:     REGISTRATION NUMBER: 33,401
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 415 742-7356
:       TELEFAX: 415 742-7217
:
: INFORMATION FOR SEQ ID NO: 12:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 233 amino acids
:     TYPE: amino acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:     MOLECULE TYPE: protein
:
: US-08-032-848C-12
:
: Query Match 20.6%; Score 254; DB 1; Length 233;
: Best Local Similarity 32.0%; Pred. No. 1,le-17;
: Matches 73; Conservative 39; Mismatches 86; Indels 30; Gaps 10;
:
: QY 15 RYRVNNVWGAETAQ-----CIEVGETGNFTTRADHDGNNVAAYPAIFYGCH 64
:   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
: Db 16 KYLENNNWGMDDEIKMGQOTIFYNSPISMG---NMWHMPSSTH-----SKATPSLYSGCH 68
:
: QY 65 WGA-CTSNGLPRRYOELSDVRTSWTLPLPTTGRNNAADYDWFSPVTVNSGNGSGGAELM 123
:   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
: Db 69 WTAGTENSGLPIQISSNKSITSNTYISIKATGYNAADYDWFHTTDDANMDSPTDELM 128
:
: QY 124 IWIN-WNGGVMPGSGRVATVELAGATWEY---WYADWD---WNIAVARTPTSYSPL 175
:   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
: Db 129 IWLNDTNAG--PAGDYIETVFLGDSSMNVEFKMINNADNGGWNVFSFVHISGTSNAS-L 185
:
: QY 176 DLKAEIDAV-ARGYIRPEWYLAHAYETGELMEGGAGLRSAFSTVQ 222
:   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
: Db 186 NIRHPTDYLVQTKWMNDEKYYSSVEFGHEIRGGQGDIDTEMRYDVK 233
:
: RESULT 9
: US-09-216-295-24
: : Sequence 24, Application US/09216295
: : Patent No. 6268328
: : GENERAL INFORMATION:
: :   APPLICANT: Mitchinson, Colin
: :   APPLICANT: Wendt, Dan J.
: :   TITLE OF INVENTION: No. 6268328e1 Variant EgIII-Like Cellulase Compositions

```

```

; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Erwinia carotovara
; US-09-216-295-24

Query Match          19.9%; Score 246; DB 4; Length 263;
Best Local Similarity 31.6%; Pred. No. 8,1e-17;
Matches 72; Conservative 40; Mismatches 84; Indels 32; Gaps 11;

```

[illegible]

```

RESULT      10
US-09-216-295-5
; Sequence 5, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Mendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-09-216-295-5

Query Match          18.2%; Score 224.5; DB 4; Length 259;
Best Local Similarity 29.6%; Pred. No. 1,1e-14;
Matches 76; Conservative 36; Mismatches 92; Indels 53; Gaps 12;

Oy      1 ELGGWDARDVAGGRRYINNVWGAET--AOQIV--GLEGNFTIRADHDNG--NNVA 54
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      20 OLCDQY--ATTGGGYTTINNIMKGDASSGSCCTTVNASSASGISTWKMNSSGENSVK 77
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      55 AYPAYIFCGHMGACTSNSGL--PRRVOELSDVRTSWTLTPITTG--RNAAVDIWFSEVT 110
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      78 SY-----ANSGLFENKKLVISIQIPTTARMSVDNTGIRADVAYDLETTADI 124
        :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      111 NSGANCSGCAELMTVLANNNGVMPCGSKRAYVELAGATWEWYV--ADMWNVIATKRTPPT 169
        :|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      125 NHHVT-WSGDYELMTIWLARYGGVOPIGSQIATAATVDGTOMELTYGANGSOKTYSFYAPPI 183
        :|| | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy      170 TSYSELDELKAFID-----DAVRAGRIR-----PEWLLHAVENTGE 204
        || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      184 TSF--OGDVNDFFKYITONHGFPASSQYLTILDTQFGTEPPTGCPALTSVNMSASVOAQCFE 242

```

QY 205 LMEGAGLRSDPSVTY 221
Db 243 PMONGAGLAVNSFSSTV 259

RESULT 11

US-09-216-295-7
Sequence 7, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328e1 variant EGIII-like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 239
TYPE: PRT
ORGANISM: Aspergillus kawachii (2)
US-09-216-295-7

Query Match 17.3%; Score 214; DB 4; Length 239;
Best Local Similarity 32.1%; Pred. No. 1.1e-13;
Matches 78; Conservative 37; Mismatches 84; Indels 44; Gaps 15;

QY 1 ELGCRMDARVAGGRYVNNVMAET---AQCIEV--GLETGFTITRADHDNG--NNVA 54
Db 20 QLCQY--ATYTGCVYIINNLMGKDGSGSQCTTVNSASSAGTSWSTKMNWSGGENSVK 77
QY 55 AYPATYFGCHWGACTSNSGL---PRVQELSDV---RTSWTLPTITGRNNAAYDIWFS 107
Db 78 SY-----ANSGLSFNKKIVSQISHPITARMSYDNTGICIRGR---AYDLETA 121
QY 108 PVTNSGNGYSGAEIMLWLMNNGVMPGSGRVATVELAGATWEVY---ADMWNYIAY 163
Db 122 ADIHVW--WSGDYELMIMLWLRYGCVQPIGSIATATVETGQWELMVGNGAKTYSFVA- 179
QY 164 RRTPTTSVSLDKAFIDDAVA--RGYIRPEWYL--HAVETGFELMEGG--AGLRSDPSV 219
Db 180 --ANPITSF--QGDINDFFKYLTONHGFPASSQYLITLALDPTGTEPFGPATLVADMSA 236
QY 220 TVQ 222
Db 237 SVQ 239

RESULT 12

US-08-032-848C-13
Sequence 13, Application US/08032848C
Patent No. 5475101
GENERAL INFORMATION:
APPLICANT: Waid, Michael
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Welss, Geoffrey L.
APPLICANT: Larens, Edward
APPLICANT: Lorch, Jeffrey D.
TITLE OF INVENTION: Purification and Molecular Cloning of
TITLE OF INVENTION: EG III Cellulase
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/032,848C
FILING DATE: MAR 17 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 742-7356
TELEFAX: 415 742-7217
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 221 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-032-848C-13

Query Match 17.3%; Score 213; DB 1; Length 221;
Best Local Similarity 31.1%; Pred. No. 1.2e-13;
Matches 73; Conservative 40; Mismatches 92; Indels 30; Gaps 12;

QY 1 ELGCRMDARVAGGRYVNNVMAET---AQCIEV--GLETGFTITRADHDNG--NNVA 54
Db 4 QLCQY--ATYTGCVYIINNLMGKDGSGSQCTTVNSASSAGTSWSTKMNWSGGENSVK 61
QY 55 AYPATYFGCHWGACTSNSGL---PRVQELSDVTSWTLPTITG--RWNAAYDIWFSPT 110
Db 62 SY-----ANSGLTFNKKIVSQISHPITARMSYDNTGIRADVAAYDLETAADI 108
QY 111 NSGNGYSGAEIMLWLMNNGVMPGSGRVATVELAGATWEVY--ADMWNYIAYRRTPT 169
Db 109 NHVT--WSGDYELMIMLWLRYGCVQPIGSIATATVETGQWELMVGNGAKTYSFVA 167
QY 170 TSVSLDLKAFIDDAVA--RGYIRPEWYLHAVETGFELMEGG--AGLRSDPSVTYQ 222
Db 168 TSF--QGDVNDFFKYLTONHGFPASSQYLITLQFTEPFGPATLVADMSA 221

RESULT 13

US-09-216-295-10
Sequence 10, Application US/09216295
Patent No. 6268328
GENERAL INFORMATION:
APPLICANT: Mitchinson, Colin
APPLICANT: Wendt, Dan J.
TITLE OF INVENTION: No. 6268328e1 variant EGIII-like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT FILING DATE: 1998-12-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 253
TYPE: PRT
ORGANISM: Humicola insolens
US-09-216-295-10

Query Match 17.2%; Score 212; DB 4; Length 253;
Best Local Similarity 30.9%; Pred. No. 1.8e-13;
Matches 71; Conservative 35; Mismatches 94; Indels 30; Gaps 12;

QY 1 ELGCRMDARVAGGRYVNNVMAET---AQCIEV--GLETGFTITRADHDNG--NNV 53
Db 37 ELGYW-----SGNGYELMIMLWLRYGCVQPIGSIATATVETGQWELMVGNGAKTYSFVA 91
QY 54 AYPATYFGCHWGACTSNSGLPRVQELSDVRT--SWTLPTITGRNNAAYDIWFSPTN 111
Db 111

Db 92 KSYPYGKQIQRG-----RKISDINSMTSVSWTYDR-TDIRANVAVDYFTADPD 141
QY 112 SGNYSGGAELMTIMLNMGVMPGSRVATVELAGATWEVYADMDNNTIAYRRTPTTS 171
Db 142 HPN-WGDDYELMTLWLRARYGGIYPIGTFRHSOVNLAKRTWDLA-TGYNNGMRYSELPSPGD 199
QY 172 VSEL--DLKAFIDDAVA-RGYIRPEWYLHAVETGFELMEG-AGLRADF 217
Db 200 IRDFSCDIKDFNFYLERNHGYPAREQNLIVYQVTECFGTGPARFCRDF 249

RESULT 14

US-09-216-295-20
; Sequence 20, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGII-like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Emeritella desertoru
US-09-216-295-20

Query Match 17.1%; Score 210.5; DB 4; Length 246;
Best Local Similarity 28.1%; Pred. No. 2.5e-13;
Matches 71; Conservative 33; Mismatches 80; Indels 69; Gaps 12;

QY 1 ELGRMDARDVAGGRYVINNNGAETACIEVLEGTGNTTIRADHDGNNVAAYPAIY 60
Db 30 DFCGMDTATV--GNFIYNNLMGQDNA--DSGSQTG-----VDSANGNSIS----- 72
QY 61 FCGHWGACTSNSG-----LPRRYOELSDVRTSWTLPIITGRW----- 98
Db 73 ----WHTTWSGSGSSSVKRYANAAOFTSTKLNSLSIPTSW-----KNQYSTTDI 120
QY 99 --NAAVDIMFSPYTNNGNGYSGGAELMTIMLNMGVMP---GGSRVATVELAGATWEVY 153
Db 121 VAVVAYDLF---TSSAGGDSEYEIMIMIALGAGPISSTGSIATVTLGGVTWLSYS 176
QY 154 AD---WDMYIAYRRTPTTSVELDKAFIDDAVARGYIRPEWYLHAVETGFELMEG- 208
Db 177 GPNGSQOVYSEVA---SSTTESFADLMDFINYLAENGLSSQYIITHVOAGTEPPTGT 232
QY 209 GAGLRADFESVTV 221
Db 233 DATLTVSVSYSVSY 245

RESULT 15

US-09-216-295-9
; Sequence 9, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGII-like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Humicola grisei

US-09-216-295-9

Query Match 16.9%; Score 209; DB 4; Length 253;
Best Local Similarity 28.5%; Pred. No. 3.6e-13;
Matches 68; Conservative 31; Mismatches 92; Indels 48; Gaps 11;

QY 1 ELGRMDARDVAGGRYVINNNGAETACIEVLEGTGNTTIRADHDGNNVAAYPAIY 60
Db 37 ELGYW-----SCNGYELNLNLMGKDJA-----TSGMCTYLDGTNNG----- 74
QY 61 FCGHWGACTSNSGLP-----RRVOELSDVRT--SWTLPIITGRWNAAY 102
Db 75 -GIOWMTAMEMOGAPDNVKNYPYVGKQIQRGKISDINSMTSVSWTYDR-TDIRANVAV 132
QY 103 DIMFSPYTNNGNGYSGGAELMTIMLNMGVMPGSRVATVELAGATWEVYADMDNNTIAY 162
Db 133 DVEFTADPDHPN-WGDDYELMTLWLRARYGGIYPIGTFRHSOVNLAKRTWDLA-TGYNNGMRY 190
QY 163 YRRTPTTSVSEL--DLKAFIDDAVA-RGYIRPEWYLHAVETGFELMEG-AGLRADF 217
Db 191 YSFLPSPGDIRDFSCDIKDFNFYLERNHGYPAREQNLIVYQVTECFGTGPARFCRDF 249

Search completed: August 15, 2002, 10:52:08
Job time: 172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:51:39 ; Search time 29.16 Seconds

(without alignments)
731.544 Million cell updates/sec

Title: US-10-003-759-2_COPY_40_261

Perfect score: 1234
Sequence: 1 ELGCRMDARDVAGGRYVIN.....FELMEGAGLRSADEFSVTVQ 222

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length DB	ID	Description
1	352	28.5	382	2	JC2571
2	252.5	20.5	264	2	JU0328
3	213	17.3	237	2	SI2610
4	200.5	16.2	274	2	A72241
5	194	15.7	151	2	H70895
6	163.5	13.2	239	2	H55831
7	153	12.4	258	2	H72240
8	142.5	11.5	332	2	G90291
9	139	11.3	334	2	G90360
10	95.5	7.7	322	2	H90425
11	93.5	7.6	673	2	B70528
12	93	7.5	741	2	G69514
13	91.5	7.4	690	2	T27357
14	91	7.4	950	2	B87611
15	88.5	7.2	261	1	SI2745
16	87	7.1	604	2	B84221
17	87	7.1	1449	1	VGTHES
18	85	6.9	957	2	H82261
19	84.5	6.8	198	2	AH2306
20	84.5	6.8	765	2	S76795
21	84	6.8	468	2	AD0585
22	84	6.8	656	2	H84649
23	83.5	6.8	475	2	T35697
24	83	6.7	302	2	H90717
25	83	6.7	323	2	H85567
26	83	6.7	468	2	H64802
27	83	6.7	595	2	T29434
28	82.5	6.7	1090	2	S59077
29	82.5	6.7	1449	1	A43573

30	81.5	6.6	1447	1	VGTHE2	E2 glycoprotein pr
31	81.5	6.6	1447	1	VGTHE3	E2 glycoprotein pr
32	81	6.6	474	1	G72658	probable Vir B11 A
33	80.5	6.5	318	1	SI6016	N-acetylmuramoyl-L
34	80.5	6.5	327	2	S34434	antigen 85-B - Myc
35	80.5	6.5	370	2	S75625	hypothetical prote
36	80.5	6.5	520	2	F91080	hypothetical prote
37	80.5	6.5	520	2	G85925	hypothetical prote
38	80	6.5	316	1	A42936	N-acetylmuramoyl-L
39	80	6.5	1374	2	AE3259	extracellular seri
40	79.5	6.4	244	2	S23099	lectin I, anti-HO
41	79.5	6.4	453	2	B83380	hypothetical prote
42	79.5	6.4	700	2	F64897	probable tonB-depe
43	79.5	6.4	729	2	T35028	probable glycosyl
44	79.5	6.4	837	1	A56681	penicillin amidase
45	79.5	6.4	986	2	SI2021	thermoactive cellu

ALIGNMENTS

RESULT 1
JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Streptomyces rochei
C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 22-Oct-1999
C:Accession: JC2571; S34392
R:Perito, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastromeli, G.
Gene 148, 119-124, 1994
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A:Reference number: JC2571; M01D:95011642
A:Accession: JC2571
A:Molecule type: DNA
A:Residues: 1-382 <PER>
A:Cross-references: EMBL:X73953; NID:g9393391; PIDN:CA52139.1; PID:g9393392
A:Note: this cellulolytic strain was isolated from the gut of termites
C:Genetics:
A:Gene: eg1S
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
C:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-37/Domain: signal sequence #status predicted <SIG>
F:38-382/Product: endoglucanase #status predicted <MAT>
F:279-380/Domain: bacterial cellulose-binding domain homology <BCB>
F:280-379/Disulfide bonds: #status predicted

Query Match 28.5%; Score 352; DB 2; Length 382;
Best Local Similarity 36.2%; Pred. No. 2e-22;
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

QY	11	VAGGRYVINNNWGAETACIEVLEGTNTTIRADHDNGNNA--AYRAIYGCCHGAC	68
DB	49	VIOGRYVONNNRMTSATQCV-TATDSG-FRVYQADSVPTNGAPKYPGVFGCHATNC	106
QY	69	TSNSGLPERRVQELSDVTSWTLPITGRNNAADVIFSPVTSNGNGYSGCAELMIILNW	128
DB	107	SPGALPARISIGSSAPSSISYGFVDNAVYNASDYIDLDPTRTDG--VNRTEIMWFNR	164
QY	129	NGVMPGSRVATVEAGATWEVYVADMDNVIAYRRRTPTTSVSELDLKAFLDDAVARG	188
DB	165	VGOIQIPRISQVGTASVAGRTVEWVSGNGTNDVLSFVAPSAMSSMSFDVMDFVATVARG	224
QY	189	XIREPWYTHAVENFELMEGAGLRSADEFSVTY 221	
DB	225	LAGNDWYLTISIQAFEPWONGAGLAVNSFSSTV 257	

RESULT 2
JU0328

cellulase (EC 3.2.1.4) precursor - Erwinia carotovora subsp. carotovora
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Erwinia carotovora subsp. carotovora
C>Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 08-Oct-1999
C:Accession: J00328
R:Saarilahti, H.T.; Henriksat, B.; Palva, E.T.
Gene 90, 9-14, 1990
A:Title: Cells: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora
A:Reference number: J00328; M0ID:90337352
A:Accession: J00328
A:Molecule type: DNA
A:Residues: 1-264 <SAA>
A:Cross-references: GB:M32399; NID:q148389; PIDN:AAA24817.1; PID:q148390
A:Experimental source: strain SCC3193
A>Note: part of this sequence, including the amino end of the mature protein, was comfili
C:Genetics:
A:Gene: cels
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-33/Domain: signal sequence #status predicted <Sig>
F:33-264/Product: cellulase #status experimental <Mat>

Query Match 20.5%; Score 252.5; DB 2; Length 264;
Best Local Similarity 31.6%; Pred. No. 3.8e-14;
Matches 72; Conservative 39; Mismatches 86; Indels 31; Gaps 10;

QY 15 RYRVNNWGAETAO-----CIEVLENGFTITRADHNGNNVAAYRFGCH 64
DB 48 KYLENNWAKDIDKMOQITFYNSPISWG---WNHWPSTH---SVAAVPSLVGWH 100
QY 65 WGA-CTNSGSLPRVVOELSDVTSWTLPTTGRNNAAYDIMEVSPYNSGNGSGAELM 123
DB 101 WTQGYENSLPQLSSNSMTSNVYTSIKATGYNAAYIMHTDKAMWDSPTDELM 160
QY 124 TWLN-WNGCVMGGSRAVAYELAGATWEWYADM-----DNNYAYRRTPTTSVSEL 175
DB 161 IWLNDTNAG--PAGDYIEFYFLGDSSWNV-FKGMINADNGGNNVFSFVHTSGTNSAS-L 216
QY 176 DLKAFIDDAV-ARGYIRPEWYLAHVEGPELMEGAGLRSDASVTVQ 222
DB 217 NHHFTDLYVOTKQMSDEKTVISSEFGEIEFGDGOIDTEWRVDVK 264

RESULT 3
S12610
cellulase (EC 3.2.1.4) precursor - Aspergillus aculeatus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Aspergillus aculeatus
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C:Accession: S12610; S14118; S40186; J00458
R:Ooi, T.; Shimmyo, A.; Okada, H.; Murao, S.; Kawaguchi, T.; Arai, M.
Nucleic Acids Res. 18, 5884, 1990
A:Title: Complete nucleotide sequence of a gene coding for Aspergillus aculeatus cellula
A:Reference number: S12610; M0ID:91016934
A:Accession: S12610
A:Molecule type: DNA
A:Residues: 1-237 <COO1>
R:Ooi, T.; Shimmyo, A.; Okada, H.; Hara, S.; Ikenaka, T.; Murao, S.; Arai, M.
Curr. Genet. 18, 217-222, 1990
A:Title: Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase) from Aspergill
A:Reference number: S14118; M0ID:91064758
A:Accession: S14118
A:Molecule type: mRNA
A:Residues: 1-237 <COO1>
A:Cross-References: EMBL:X52525; NID:g2287; PIDN:CAA36757.1; PID:g2288
A:Accession: S40186
A:Molecule type: protein
A:Residues: 17-18;42-49, 'X', 51-54, 'X', 66-79;90-111,136-205, 'XX', 208-211 <OO13>
C:Genetics:

A:introns: 138/2; 212/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; pyroglutamic acid
F:1-16/Domain: signal sequence #status predicted <Sig>
F:17-237/Product: cellulase #status experimental <Mat>
F:17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experi

Query Match 17.3%; Score 213; DB 2; Length 237;
Best Local Similarity 31.1%; Pred. No. 7.6e-11;
Matches 73; Conservative 40; Mismatches 92; Indels 30; Gaps 12;

QY 1 ELGRMDARDVAGRRYRVINWGAET--AOCTEV--GLETGNFTITRADHNG--NNVA 54
DB 20 QLCDQY--ATYTGQVYITNNLNGKADAGSGCCTVNASASACTSMSTKMNWGSNSVY 77
QY 55 AYPAIRFGCHWAGCTNSGL---PRVQELSDVTSWTLPTTGG--RWNAAYDIMEVSPY 110
DB 78 SY-----ANSGLTFENKLVISOIPTTARMSYDWTGIRADVADLFTAAADI 124
QY 111 NSGNGSGAELMIMLNMGCVMPGSRVATVELAGATWEWY--ADMDNVIYRRTPTT 169
DB 125 NHYT-WSGDELMIMLARIGGVOPISQIATATVDCOTELMIGANGSOKTYSFVAFTPI 183
QY 170 TSVSELDLKAFLIDAVA-RCYIRPEWYLAHVEGPELMEGCG-AGLRSDASVTVQ 222
DB 184 TSE-QGDVNDFFKYLQNHGFPAASQYLITLQGTPEFTGPGATLVSMSASVQ 237

RESULT 4
A72241
endoglucanase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72241
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; M0ID:99287316
A:Accession: A72241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <ARN>
A:Cross-references: GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AAD36592.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TW1525

Query Match 16.2%; Score 200.5; DB 2; Length 274;
Best Local Similarity 27.9%; Pred. No. 1e-09;
Matches 70; Conservative 39; Mismatches 85; Indels 57; Gaps 15;

QY 6 WDARDVAGRRYRVINWGAETAOCTIEVLENGFTITRADHNGNNVAAYRFGCH- 64
DB 44 WNKSYEG-----ETWLKFDEKEVFAADLVNLYQND-----SMVHGYPEITYGYKP 92
QY 65 WGACTNSG---LPRVVOELSD--VRTSWTL-----PTTGRNNAAYDIMEVSPYNSGN 114
DB 93 WAG--HNSGVEFLPVKVKDLPDFYVLDDYSIWEENMLPI-----NLAETWITRSPDQTS 145
QY 115 GYSGAELMIMLNMGCVMPGSRV---ATVELAG-----ATWEWYADMDNVIYRRT 166
DB 146 VSSGDAELINWF--YNNVLMFGGOKVDEFTTYEINGVKQETKMDVFAFGMGWYLAFLRL 204
QY 167 TPTTSVS-ELDLKAFIDDA---VARGYIRPEWYLAHVEGPELMEGCA----- 210
DB 205 TPKEBEKVKINVDFOKAAEVYKKHSTRD---NEELYFCWELGTGEGDPNTAAKF 261

Qy	211	GLRSADFSVTV	221
Db	262	GWTFRDFSVEV	272

RESULT 5
H70895
hypothetical protein RV1090 - Mycobacterium tuberculosis (strain H37Rv)
..Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence, revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: H70895
R/Coile, S.T., Brosch, R., Parkhill, J., Garner, T., Churcher, C., Harris, D., Gordon, S., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Rastredrum, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R., Hulstov, J.E., Taylor, B.G., Whitehead, S., Barrell, B.G., et al. Deciphering the tubercle bacillus genome

Query Match Similarity	15.7%	Score	194;	DB 2;	Length	151;			
Best Local Similarity	33.3%	Pred. No.	1.9e-09;						
Matches	51;	Conservative	16;	Mismatches	78;	Indels	8;	Gaps	4;

Qy 7/2 SGIPRRVQELSDVIRISWTLFBITTTGRNNAAYDWF--SPVTNSGNGYSGGAEIMITWLNWN 129
: || | :: ||| |::||| : ||| |
Db 3 TNLPTFVGQILSAPISIDYNYPTTGWDASDYDCLDSTFKTTGVN---QQEIMITWLNHQ 58

59 GSIQPVGSPVGNITIEGKNFYVWDGSGMNNMAYVATEP-IEVWSFDVMSFVDHTATME 117

QY 189 YRPEWYLHAVETGFELMEGAGLRSADESVTV 221
 Db 118 PITDSWYLTIRAGLEPWSGDSVGLGVDSFSAKV 150

RESULT 6
S55931
cellulase (EC 3.2.1.4) precursor - *Aspergillus niger* -

Query Match	13.28;	Score 163.5;	DB 2;	Length 239;
Best Local Similarity	27.38;	Pred. No. 1.2e-06;		
Matches 63;	Conservative 38;	Mismatches 97;	Indels 33;	Gaps 11;

```

QY      2  LCGHMDADVAGGRYRYIANNWG---AETACIEVG---LENGFTTIRADHDNG-NNVA 55
      19  MCSOYDS-ASSPPY SVYNQNLMEGYOCTGSOCCYVDKLSSGASWHKWTWMSGEGEYVKS 76
Db      56  YPAIFYCGCHWGACTSNSGL---PRVOELSDVRTSWTLTPIPTTG-RNNAAYDIWFSPVTN 111
      77  Y-----SNSGLTFDRKLTLYSDVSISPIPSVWMSQDDTIVQADVSYDL-FTANA 122
QY      112  SGNQYSGCAELMILNNGCYMPCGSKRYAYVELGATIEWY----ADMDMNTIAYRT 166
Db      123  DHATSSGGEYELMILARYGSVQPIGKQIATFYATGKSMIEWYGTSTOGAEGEOKTYSFVAG 182
QY      167  TPTTSVSELDKAFIDVAVA-RGYIRPEWYTHAAETGSELMGEGAGLRSD 216
Db      183  SPISMWSG-DIKDFEFLITQWOGCPASSQHLLITLQCTGEPTTGPAFTYD 232

```

RESULT 7
H72240
endoglucanase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72240
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: H72240
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <ARN>
A:Cross-references: GB:AE001800; GB:AE000512; NID:g49682090; PIDN:AAD36591.1; PID:g4968
A:Experimental source: strain MSB8
C:Genetics:
;Gene: TML524

Query Match	12.4%;	Score 153;	DB 2;	Length 258;
Best Local Similarity	27.7%;	Pred. No. 1e-05;		
Matches 51;	Conservative 24;	Mismatches 87;	Indels 22;	Gaps 8;

Qy	53	VAAPAIYEGH-GNACTS-NSGJPRRQELTS--DVRTSMILPITIGRNAAADIESP	108
Db	63	VLGIPETTYIKRPENHTAEGSKLPVYSSMKSFSEVSEFDIHHPSLPLNFAMETWLR	122
Qy	109	VTNSGNGSGGAEILMILMNGVMPGGSR-----VATVELAGATWEVYADMDNY	160
Db	123	EKYQTEASIGVEILVMVFYFN-NLTPGGEKIEEPTIPVLNGESVEGMEILMAEMQDY	181
Qy	161	IAYRTPTTSVS-ELDLKAFIDA-----VANGIYRPEYLAHAVETGPELMGGAGLR	213
Db	182	LAFRLKDPVKKGRVYKFDVRRFLDAAGKALSSASAVKDFEDLYTFWVEIGTEF--GSPETK	239
Qy	214	SADF	217
Db	240	SAQF	243

```

RESULT      8
G90291
endoglucanase precursor [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: G90291
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.U.; Chong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder J., R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

```

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90291

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-332 <KUR>

A:Cross-references: GB:AE006641; NID:g13814564; PIDN:AAK41590.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO1354

Query Match

Best local similarity 11.5%; Score 142.5; DB 2; Length 332;
Matches 60; Conservative 32; Mismatches 76; Indels 75; Gaps 16;

6 WDARDAAGGKRYINNVWAGETACIEVGLTGNFTITRA---DHDNGNNVAAYPAIYFG 62

100 WNAKNTNGNYTWENPL-----ARTLSV-----SFNLTQVKPLEWNG-----YPEIYVG 144

63 CH-WGACTSNSGLPRRVOELSDVTSWTLT-----PITGGNNAAVDIMESPVTN 111

145 RKWMDIAYAGNIPPMKIGNMTPEWVSFYILTKLDPISINFDIASDMAIYRPOIAFSPGTA 204

112 SGNGYSGAELMILNMWNGVMPGSGRVATV-----ELAGATWEVW---YADW-DWN 159

205 PGN---GDIEIHWML-FSONLQPAAGEQGVVPIYINHTLVNATQVWEMKGVPMGWE 260

160 YIAYRRTPPTSSELDLKAIFIDA--VARGYI--RPEWHLAVE-----TGFEL--W 206

261 YIAFR-----PDGKKVTNGYVSYPENFLKALNSFTSYNTITNYLTDM 303

207 EGG 209

304 EFG 306

RESULT 9

endoglucanase precursor [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: G90360

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F

aretz, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90360

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-334 <KUR>

A:Cross-references: GB:AE006641; NID:g13815224; PIDN:AAK42142.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO1949

Query Match 11.3%; Score 139; DB 2; Length 334;

Best local similarity 24.7%; Pred. No. 0.00021;

Matches 60; Conservative 26; Mismatches 69; Indels 88; Gaps 16;

21 NWGASTAOCIEVGLTGNFT---ITRA-----DHDNGNNVAAYPAIYFG 62

100 NMMNAKT-----WNGNTWVFNPLTRILSYFNLTQVNPLOWNG-----YPEIYVG 146

63 CH-WGACTSNSGLPRRVOELSDVTSWTLT-----PITGGNNAAVDIMESPVTN 111

147 RKWMDIAYAGNIPPMKIGNMTPEWVSFYILTKLDPISINFDIASDMAIYRPOIAFSPGTA 206

112 SGNGYSGAELMILNMWNGVMPGSGRVATV-----ELAGATWEVW---YADW-DWN 159

207 PGN---GDIEIHWML-FSONLQPAAGEQGVVPIYINHTLVNATQVWEMKGVPMGWE 262

160 YIAYRRTPPTSSELDLKAIFIDA--VARGYI--RPEWHLAVE-----TGFEL--W 206

261 YIAFR-----PDGKKVTNGYVSYPENFLKALNSFTSYNTITNYLTDM 303

207 EGG 209

304 EFG 306

RESULT 10

hypothetical protein celb [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: H90425

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

aretz, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: H90425

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-673 <COU>

A:Cross-references: GB:Z96796; GB:AL123456; NID:g3261797; PIDN:CAB09571.1; PID:g32617

C:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV0457C

Query Match 7.7%; Score 95.5; DB 2; Length 322;

Best local similarity 20.3%; Pred. No. 1;

Matches 44; Conservative 33; Mismatches 95; Indels 45; Gaps 8;

33 VGLTGNFTIT-----RADHDNGNNVAAYPAIYFG--CHWGACTSNS----- 72

83 IGYALGNVMTITINYLHVAIINLSQISKISSNVVDYPGIMGOELMPEWRYRTLOPL 142

73 GUPRRVOELSDVRT--SWTLTPITGGNNAAVDIMESPTNSGNGYSGAELMILNMWNG 130

143 SLPMIYLRPLNPFISILNYSYILNGSINDSFSDIWLSONPNITSLQGFPEIMIMYWN 202

131 G-----VMPGSGRVATV---ELAGATWEVWVA-----DWDNMYIAYRRTPPTSVS 173

203 NLSHTPYFIYVGNMSIPTILNGKIEINLSMEVYVLPRTGSANGWTGYFSLPKPRAERG 262

174 E-----LDIKAFIDAIVARGYIRPEWYIHAVETGE 204

263 VPIGYILKNMSYIERKAGVINYVNTYIDAIQVGM 299

RESULT 11

probable peptidase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70528

R:Conor, R.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: B70528

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-673 <COU>

A:Cross-references: GB:Z96796; GB:AL123456; NID:g3261797; PIDN:CAB09571.1; PID:g32617

C:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV0457C

C:Superfamily: prolyl oligopeptidase

Query Match
Best Local Similarity 22.6%; Score 93.5; DB 2; Length 673;
Matches 56; Conservative 14; Mismatches 61; Indels 117; Gaps 12;

QY 4 GRWDARVAGGRYRINNVGAETACIEVGLTETGNTFTTRADHDGNNVAAYPAITFG 62
DB 467 GRLMLAR---GGFYALANIRGGGE-----YG 489
QY 63 CHNGACSNGLPRRVQELSDVRSWTLPTTGRMAADIVFSPYNSGNGYSGAEL 122
DB 490 PGWHTQAMREGROKVDFAVAADLVTRGITTA-----DOLGARGSS-- 532
QY 123 MIMLWNGVMPG-----GSRVATVE-----LAGATWEVWYAD---WDW 158
DB 533 -----NGGLMGIMLTGYPEKFGALYCDVPPLDMKRYHLLLAGASMAEYGDNDNDW 586
QY 159 NYI-----AYRRTPP---TTSVSELDLKAFTIDAVARGYR-----PEW 194
DB 587 KFTISESPYONISANRKYPPVLMTTSTR-----DDRYPHGHARFMTAALQAAHGPVW 638
QY 195 YLHAVENG 202
DB 639 YVENIEGG 646

RESULT 12

hypothetical protein AF2119 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Nov-1999
C:Accession: G69514
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kalne, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: G69514
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-741 <RLE>
A:Cross-references: GB:AE000958; GB:AE000782; NID:g2689281; PIDN:AA89151.1; PID:g264842
C:Superfamily: Archaeoglobus fulgidus hypothetical protein AF2119

Query Match
Best Local Similarity 24.4%; Score 93; DB 2; Length 741;
Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

QY 36 ETGNTFTTRADHDGNNVAAYPA-----TYFGCHMGACTSNSG-----LPRR-- 77
DB 204 DTTTITLIARVNSGNVPPDKVFAEFDSPQRIETG--GAWTNSSGIAKLSPFANVG 260
QY 78 -----VOELSDVRSWTLPTTGRMAADIVFSP---VTNSGN-GYSGAELM 123
DB 261 LSKLRVNFYAKIEDVWTCNCAATTTNRAILANVAITGCSYDITLVGRMYSGAD-W 319
QY 124 IWLNV-----NGVMP-----GSSRVATVELAGATWEVWY----- 153
DB 320 VRNMYVDENSVLYKPYPIVVOKEFTGNR-----ASVTIMKYGLDDYCTDPNCHREGI 371
QY 154 -----ADWDMNTIAYRRTP 168
DB 372 YGNFDDADMDGACIAGVSTP 392

RESULT 13

T27357

hypothetical protein Y70G10A.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27357
R:Uday, C.
submitted to the EMBL Data Library, October 1998
A:Reference number: T20354
A:Accession: T27357
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-690 <NID>
A:Cross-references: EMBL:AL032660; PIDN:CAA21751.1; GSPDB:GN00021; CESP:Y70G10A.3
A:Experimental source: clone Y70G10A
A:Gene: CESP:Y70G10A.3
A:Map position: 3
A:Introns: 61/3; 84/2; 185/1; 250/2; 326/3; 375/1; 398/3; 439/2; 490/3; 628/1; 655/1

Query Match
Best Local Similarity 20.5%; Score 91.5; DB 2; Length 690;
Matches 42; Conservative 28; Mismatches 62; Indels 73; Gaps 8;

QY 63 CHNGACTSN-----SGLPVRQVE-----LSDVRSWTLPTTGRW 98
DB 31 CGYGACTPSWLOGFHNACKLLVVLGICAFIOFVNAIFPVGLSTERRRKMSTHTGII 90
QY 99 NAAVD-----IMFSPVTSNG-NGYSGAELMIMLWNGVMPGSGRVATV----- 142
DB 91 SSWDFAVLLVFPVPCWGNNGHGR-----WIGMGVIMALSLCALPHMWDIYHPD 145
QY 143 -----ELAGATWEVWYADWDMNTIAYR-----TPTTSVSELDL 178
DB 146 VNDLTNQTDFGQACNRDPCAGKPHSSMFNPYFWMFILQTLHGVSPTLFSIG----T 201
QY 179 AFIDDAVARGYIREWYLAHVENGF 203
DB 202 TYMDENVSQKASPYIAIHAVALTSF 226

RESULT 14

Toob-dependent receptor [Imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87611
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87611
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-950 <STO>
A:Cross-references: GB:AE005673; NID:g13424546; PIDN:AAK24866.1; GSPDB:GN00148
C:Gene: CC2924

Query Match
Best Local Similarity 22.0%; Score 91; DB 2; Length 950;
Matches 54; Conservative 32; Mismatches 96; Indels 64; Gaps 14;

QY 2 LCGRWDAADVAAGGRYRINNVGAETACIEVGLTETGNTFTTRADH----- 47
DB 434 LTGEWVSD---KLRYDGH-----IGQETSDYDIPISDKRYTAFGGLTIDR 478
QY 48 -DNGNNVAAYPAITFGCHMGACTSNSGLPRRVQELSDVRSWTLPTTGRMAADIVF 106
DB 479 GDGSKNT-----YKMTTIDANN---YRAHET-DFSATYQTTELKKNFNAADVFS 525

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:58:02 ; Search time 15.93 Seconds

(without alignments)
539.594 Million cell updates/sec

Title: US-10-003-759-2_COPY_40_261

Perfect score: 1234
Sequence: 1 ELGGRMDARDVAGGRYRIN.....ELMEGAGLRSDPSVTWQ 222

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	252.5	20.5	264	1	GUNS_ERMCA
2	213	17.3	237	1	GUN_ASPAC
3	163.5	13.2	239	1	GUNA_ASPAK
4	93	7.5	625	1	BGAL_LACSK
5	93	7.5	741	1	VL19_ARCFU
6	88.5	7.2	261	1	XYNA_CLOSA
7	87	7.1	1449	1	VG12_CVPEF
8	86.5	7.0	269	1	EL2_BOVIN
9	83.5	6.8	475	1	ABFE_SMRCO
10	83	6.7	468	1	YBFK_ECOLI
11	82.5	6.7	1090	1	GUXB_CELFI
12	82.5	6.7	1449	1	VG12_CVPMI
13	81.5	6.6	225	1	THTR_CORGL
14	81.5	6.6	1447	1	VG12_CVPEP
15	81.5	6.6	1447	1	VG12_CVPPU
16	80.5	6.5	318	1	ALYS_BRH3
17	79.5	6.4	700	1	YNCD_ECOLI
18	79.5	6.4	986	1	GUNE_CLOSR
19	79	6.4	475	1	ABFB_STRLI
20	78.5	6.4	233	1	XYNB_MAGGR
21	78.5	6.4	479	1	CBSA_SULSO
22	78.5	6.4	1447	1	VG12_CVPRK
23	78.5	6.3	722	1	ACMI_DROME
24	77.5	6.3	846	1	PAC_ECOLI
25	77.5	6.3	1356	1	HERI_PODAN
26	76.5	6.2	270	1	KITM_MOUSE
27	76.5	6.2	318	1	ALYS_STRPN
28	76.5	6.2	327	1	A85B_MYCLE
29	76.5	6.2	350	1	MLFI_MALFE
30	76.5	6.2	474	1	SHU1_ECOLI
31	76	6.2	240	1	XYNC_STRLI
32	76	6.2	3396	1	POLG_DENIS
33	75.5	6.1	624	1	TSPE_BPSFV

34	75.5	6.1	641	1	MTHS_SCHPO	074927 schizosach
35	75.5	6.1	966	1	ENV_CAEVC	P31626 caprine art
36	75.5	6.1	3430	1	POLG_MNV	P06935 w genome po
37	75	6.1	441	1	GSA_PROFR	P06774 propionibac
38	75	6.1	945	1	AMPE_MOUSE	P16406 mus musculu
39	74.5	6.0	323	1	A85B_MYCBO	P12942 mycobacteri
40	74.5	6.0	461	1	PSBC_CHLRE	P10898 chlamydomon
41	74.5	6.0	513	1	GUX1_TTRRE	P00725 trichoderma
42	74.5	6.0	581	1	IRR_RAT	P04716 rattus norv
43	74.5	6.0	852	1	POL_BIVU	P03361 bovine leuk
44	74.5	6.0	872	1	GUXA_CELFI	P50401 cellulomona
45	74.5	6.0	1310	1	VAC3_HELPY	048253 helicobacte

ALIGNMENTS

RESULT	ID	GUNS_ERMCA	STANDARD	PRT	264 AA.
AC	P16630	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Endoglucanase S precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase S) (Cellulase S).				
GN	CELS.				
OS	Erwinia carotovora.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.				
OX	NCBI_TaxID=554;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=SCC3193;				
RX	MEDLINE=90337352; PubMed=2379837;				
RA	Saarihahti H.T., Henttinen B., Palva E.T.;				
RT	'Cels: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora.';				
RL	Gene 90:9-14(1990).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.				
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY H (FAMILY 12 OF GLUCOSYL HYDROLASES).				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sdb.ch/announce/ OR SEND AN EMAIL TO license@sdb-sdb.ch).				
CC	EMBL: M32399; AAA24817.1; -				
DR	PIR: J00328; J00328.				
DR	InterPro: IPR002594; Glyco_hydro_12.				
DR	Pfam: PF01670; Glyco_hydro_12; 1.				
DR	ProDom: PD004316; Glyco_hydro_12; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
FT	SIGNAL 1 32				
FT	CHAIN 33 264				
FT	SEQUENCE 264 AA: 29757 MW: E6D61388950C77AA CRC64;				

Query Match 20.5%; Score 252.5; DB 1; Length 264;

Best Local Similarity 31.6%; Pred. No. 3,1e-15;

Matches 72; Conservative 39; Mismatches 86; Indels 31; Gaps 10;

QY	15	RYRYNNWGAETAAQ-----CIEVLENGFTTTADHDNGNNVAATPAITFGCH 64
DB	48	RYRYNNWGAETAAQ-----CIEVLENGFTTTADHDNGNNVAATPAITFGCH 64
QY	65	WGA-CYNSNSGLPRVQGLSVYRTSWTLFTTTGRWNAAYIMVSPVINSNGSGGAEIM 123
DB	65	WGA-CYNSNSGLPRVQGLSVYRTSWTLFTTTGRWNAAYIMVSPVINSNGSGGAEIM 123

ID	NAME	SEQUENCE	FUNCTION	REFERENCE	REMARKS
DB	101	WAGTGTENSGLEIQTSSNKTSITSNTVTSYIKACGTGTAANVDIMFHITTDKANKMSSPTDELM			160
OY	124	ITLNL-WNGCWMPGGRSVATVELAGATWEXWYADW-----DWNTIAYRRTPTTSVSEL			175
DB	161	ITLNTTNG--PAGYIEIVFPLGDSWMN-FKGMINADNGCMWNFSFVHTSGTNSAS-L			216
OY	176	DKAFITDAV-ARGYIRPEWYLAHVETGELFEGGAGLSADFSTVQ			222
DB	217	NIRHRTDVLIVOTKQWMSDEKTYISSVEFTGFEFGGQIDITEMRYDVK			264
RESULT 2					
GUN	ASPAC	STANDARD:	PRT:	237 AA.	
AC	P22669:				
DT	01-AUG-1991	(Rel. 19, Created)			
DT	01-AUG-1991	(Rel. 19, Last sequence update)			
DT	15-DEC-1998	(Rel. 37, Last annotation update)			
DE	Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)				
DE	(Cellulase) (FI-CMCase).				
OS	Aspergillus aculeatus.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
OC	Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.				
OK	NCBI_TaxID=5033;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-F-50;				
RX	MEDLINE=91016934; PubMed=2216782;				
RA	Ooi T., Shimmyo A., Okada H., Murao S., Kawaguchi T., Arai M.;				
RT	"Complete nucleotide sequence of a gene coding for Aspergillus				
RT	aculeatus cellulase (FI-CMCase)."				
RL	Genetic Acids Res. 18:5884-5884(1990).				
RN	[2]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN-F-50;				
RX	MEDLINE=91064758; PubMed=2249253;				
RA	Ooi T., Shimmyo A., Okada H., Hara S., Ikenaka T., Murao S.,				
RA	Arai M.;				
RT	"Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase)				
RT	from Aspergillus aculeatus."				
RL	Curr. Genet. 18:217-222(1990).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- SUBCELLULAR LOCATION: Secreted.				
CC	-1- INDUCTION: BY CELLULOSIC MATERIALS AND HEMICELLULOSES.				
CC	-1- MISCELLANEOUS: WIL ALSO HYDROLYSE 1,4-LINKAGES IN BETA-D-GLUCANS				
CC	ALSO CONTAINING 1,3-LINKAGES.				
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL				
CC	HYDROLASSES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: D00546; BAA00435.1; -				
DR	EMBL: X52525; CA36757.1; -				
DR	PIR: J00458; J00458.				
DR	PIR: S12610; S12610.				
DR	PIR: S14118; S14118.				
DR	InterPro: IPR002594; Glyco_hydro_12.				
DR	Pfam: PF01670; Glyco_hydro_12; 1.				
DR	Prodom: PD004316; Glyco_hydro_12; 1.				
FW	Cellulose degradation; Hydrolyase; Glycosidase; Signal.				
KM	SIGNAL				
FT	1				
FT	16	POTENTIAL.			
FT	17	ENDOGLUCANASE I.			
FT	17	PYRROLIDONE CARBOXYLIC ACID.			
QO	MOD_RRS				
QO	SEQUENCE	237 AA; 25560 MW; 8617571A8A65931 CRC64;			

	Query Match	17.3%	Pred. 213	DB 1	Length 237
	Best Local Similarity	31.1%	Pred. No. 8.2e-12		
	Matches 73	Conservative 40	Mismatches 92	Indels 30	Gaps 12
Qy	1 ELGCGMDARDVAGGRYRINNVMAET--ADOCIEV--GLEGNFTITRADHNG--NNVA	54			
Db	20 QLCDQY--ATYTGCVYTTINNLMGKDAGSGSCCTTVNSASSAGTSMSTRKMNNGGENSEV	77			
Qy	55 AYPAITIEGCHMGACISNSGL---PRRVDLSPIKRSMTLPTTIG--RKNAAVDIWFSPYT	110			
Db	78 SY-----ANSGLTFNKKLVLSQISDIPTAAWSDONTGIRADVAADLETADI	124			
Qy	111 NSGNGYSGGAEIMLTNNNGGVPCGSEVATVELAGATWEVY--ADPMQNYIAYRRTPPT	169			
Db	125 NHHV--WQSDYELMILNARYGVQPIGSDIATATVDGQWELMYGANGSKTYSFVAPIPI	183			
Qy	170 TSVSELDLKAFIDDAVA--RGYIRPMYLIHAVETGELMBEG--AGLRASDFSTVQ	222			
Db	184 TSF--QGDVNDPEFKYLTQNHGEPASSOYLITLQFGTEPPTGGATLTSNWSMA	237			

Query Match	Best local similarity	Matches	2 LGRMWARDVAGGRYRVYNNVWG--AETQAQIEVG--LETGNGFTITRADHNG--NNVAA
13.2%	Score 163.5; DB 1;	Length 239;	
27.3%	Pred. No. 1,8e-07;		
38;	Mismatches 97;	Indels 33;	Gaps 11;

Db 19 MCSOYDS--ASSPPYVNONLMGEYOGTSGOYVVDKLSSGASWMTKWTWGSGETVKS 76
QY 56 YPAIFGCHWACTNSGL--PRRVOELSDVTSWTLPTTGG-RWNAAYDIWFSPVTN 111
Db 77 Y-----SNSGLFEDKRLVSDVSSITPSYWSODDPTNQADVSYDL-FTANA 122
QY 112 SGNGYSGGAEMLMWMNGCVMPGSRVATVELAGATWEVY----ADMDMNIAYRRT 166
Db 123 DHATSSGDYELMIMLWARGVOPIGKQIARATVYGKSMWEVWYGTSTQAGAEQKTYSEVAG 182
QY 167 TPTTSELDLKAFIDDAVA-RGIYREPMVLHVENGFELMEGAGAGRSAD 216
Db 183 SPINMSG-DIKDFPNTLTONOGPPASSQHLTLTQCCTEPTGPGATPTVD 232

RESULT 4

BGAL_LACSK STANDARD; PRT; 625 AA.
AC 048846; 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-galactosidase large subunit (EC 3.2.1.23) (Lactase).
GN LACL.
OS Lactobacillus sakei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 20017;
RX MEDLINE=96118231; PubMed=8574399;
RA Obst M., Meding E.R., Vogel R.F., Hammes W.P.;
RT "Two genes encoding the beta-galactosidase of Lactobacillus sake.";
RL Microbiology 141:3059-3066(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X82287; CAA5730.1; -
CC HSSP: P00722; 1BGI.
DR InterPro: IPR001649; Glyco_hydro_2.
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
DR Pfam: PF02837; Glyco_hydro_2_N; 1.
DR PRINTS: PRO0132; GLYDRLASE2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW Hydrolyase; Glycosidase.
FT ACT_SITE 465 465 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 533 533 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 625 AA; 72457 MW; C65BA3C46136886 CRC64;

Query Match 7.5%; Score 93; DB 1; Length 625;
Best Local Similarity 17.9%; Pred. No. 0.81;
Matches 44; Conservative 39; Mismatches 87; Indels 76; Gaps 11;

QY 14 GRKRVNNV---WGAETACIEVGLTGNFTTRADHDGNNVAAP---AIYFGCHWG 66
Db 342 GKRLVINGNHRHMHETPTTAEDEAMDIACQNRHINAIVRTSHDRLSFYNGC--- 398
QY 67 ACSNSGGLPRVQELSDVRSWTLPTTGRNN--AAVDWFSPVYNSG---NGISGGA 120

Db 399 ---DQAGIYMAETNESHSQWQMGAVPEPSWVPGSYDEMEATLDRARTNETFRKNHV 455
QY 121 ELMITMWMNGVMPGSRVATVELAGATWEVYADMDMNIAYRRTPT----- 169
Db 456 SLIFW-----SLGNSYAGSVLEKMA-----YKQDDPTLVHYEGVFA 496
QY 170 ----TSVSEL-----DLKAFIDDAVARGYIRPEMVLHAVETG-----F 203
Db 497 PEYKATISDVESHMYATPAEIKAYLDNAPQKPTLCE-YHHDGNSLIGMQSYIDLLSGY 555
QY 204 ELMEGG 209
Db 556 DMVGG 561

RESULT 5

YL19_ARCFU STANDARD; PRT; 741 AA.
AC 028161; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2119 precursor.
GN AF2119.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Krelavag A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton P.M., Spriggs T., Arlatch P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE000958; AAB89151.1; -
DR TIGR: AF2119; -
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 741 HYPOTHEICAL PROTEIN AF2119.
SQ SEQUENCE 741 AA; 82084 MW; E8C7543552231583 CRC64;

Query Match 7.5%; Score 93; DB 1; Length 741;
Best Local Similarity 24.4%; Pred. No. 0.98;
Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

QY 36 ETGNTTTRADHDGNNVAAP-----TYFGCHWACTNSG-----LPRR-- 77
Db 204 DTFITLVAKVNSGNDVPDPYKFAEFDSDGIFLG---GAMTNSGGLAKLSFIRKNNG 260
QY 78 -----VQELSDVTSWTLPTTGRNNAAYDIWFSP-----VYNSGN-GYSGAEML 123


```
Oy      16 YRVINNVWGAETAQCIEVGLTGT-NFTIRADHDNGN-----NVAAYPAIFYCGHWGAC    68
          |   |       ||| : : : : | | | |
Db      50 YSAFNTQAAPEPTITSNELGVNGGYDELMK---DYGMTSMTLKGNA-----FSICOW---    98
```

FT	CARBOHYD	405	405	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	449	449	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	516	516	N-LINKED (GLCNAC. . .)	(POTENTIAL).


```

FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 594 594 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 704 704 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 819 819 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1324 1324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1358 1358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1449 AA; 159957 MW; 971BBAE191FDD1AF CRC64;

```

Query Match Best Local Similarity 7.1%; Score 87; DB 1; Length 1449;

Matches 45; Conservative 20; Mismatches 69; Indels 62; Gaps 10;

```

OY 6 WDARDVAGRRYVNNVGAETACIEVLEGTGNTTTRADHDGNNV-AAAYATPG-- 62
DB 88 WD-----YATENSFWNHK--QRLNVVNGYPSYITVTTTRFNBSAEGALICCKGSP 137
OY 63 -----CHGCA-----CTNSGLPRRVOELSDVRSMWLTPTTGGRMAAY 102
DB 138 PTTTSSLLTCNMNGBECRLNHRKPPICPSN-----EANGCMMLY 176
OY 103 DI-WFSPVTSNGVSGGAEIMTL--NMNGVPGGSRVATVELAGATWEVYAD--WD 157
DB 177 GLQMFADAVVA---YLHGAAYRISFENMGSTVTLGDMRATILETACTIVDLMFNFVYD 233
OY 158 WNTIAYRTTPTTSVS 173
DB 234 VSYRYVNNKNGTIVVS 249

```

RESULT 8 EL2_BOVIN STANDARD; PRT; 269 AA.

```

AC Q29461;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elastase 2 precursor (EC 3.4.21.71).
GN ELA2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=98079203; PubMed=9418008;
RA Gestin M., le Huertou-Luron I., Wicker-Planquart C., le Drian G.,
RA Chail J.C., Puigserver A., Guilloteau P.;
RT "Bovine pancreatic preproelastases I and II: comparison of nucleotide
RT and amino acid sequences and tissue specific expression.";
RL Comp. Biochem. Physiol. 118B:181-187(1997).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-I-Xaa, Met-I-Xaa
CC and Phe-I-Xaa. Hydrolyses elastin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. ELASTASE SUBFAMILY.

```

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

```

DR EMBL: X97635; CAA66231.1; -
DR HSSP: P00766; 1CHG.
DR MEROPS: S01.155; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PFO0089; trypsin.1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_DOM; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; zymogen; signal.
FT SIGNAL 1 16 BY SIMILARITY.
FT PROPEP 17 28 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 29 269 ELASTASE 2.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 216 216 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 155 222 BY SIMILARITY.
FT DISULFID 186 202 BY SIMILARITY.
FT DISULFID 212 243 BY SIMILARITY.
SQ SEQUENCE 269 AA; 28856 MW; 8343B97062C67C CRC64;

```

Query Match Best Local Similarity 7.0%; Score 86.5; DB 1; Length 269;

Matches 57; Conservative 21; Mismatches 63; Indels 93; Gaps 15;

```

OY 4 GRMDARDVAGRRYVNNVGAETACIE-----VG-----LETGNFTTRAD--- 46
DB 52 GOW--RHTCGG--SLIEQNNVLTFAAHCISSSRYRVVVGQSLSIVBSGLTIAVSXVI 107
OY 47 HD-----NGNNVAY-----PAIYFGC-----HNGACISNS 72
DB 108 HEKMSNQLAOGNDIALKLASSVPLTDKIQCLPAAGTILPNVYCVYTGGRQISNG 167
OY 73 GLPRVGE-----LSDVPT-----SWTLPTTGRRNNAAYDIFSPVTSNGVSGGAEIMI 124
DB 168 ALPDIIDQGLLVVDYATCSNPSMGSFTVTKMTICAGD-----GYISSCGDSGGP----- 219
OY 125 WLMNNGVMPGSRVATVELAGATWEV---WYADDMNYIAYRTTPTTSVS 173
DB 220 LN-----CQAANROMQVHGIVGSSISLCNV--YRKPSVETPVS 256

```

RESULT 9 ABFB_STRCO STANDARD; PRT; 475 AA.

```

AC 054161;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
GN ABFB OR SCTH1.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D., Parkhill J., Barrrell B.G., Rajandream M.A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

```

CC -1 CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC arabinofuranoside residues in alpha-L-arabinosides.
CC -1 PARTWAY: XLYAN DEGRADATION.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 SIMILARITY: BELONGS TO FAMILY 62 OF GLYCOSYL HYDROLASES.
CC -1 SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC DR EMBL: AL021411; CA16189.1; -
CC InterPro: IPR000772; Ricin_B_lectin.
CC pfam: PR00652; Ricin_B_lectin; 1.
CC SMART; SM00458; Ricin; 1.
CC DR PROSITE; P550231; Ricin_B_Lectin; 1.
CC KXylan degradation; Hydrolase; Glycosidase; Signal; Lectin.
CC FT SIGNAL 1 37 POTENTIAL.
CC FT CHAIN 38 475 ALPHA-L-ARABINOFURANOSIDASE.
CC FT DOMAIN 39 166 RICIN B-TYPE LECTIN.
CC SQ SEQUENCE 475 AA; 50045 MW; 47E707EE543CA60D CRC64;

Query Match 6.8%; Score 83.5; DB 1; Length 475;
Best Local Similarity 22.5%; Pred. No. 4.1;
Matches 43; Conservative 28; Mismatches 77; Indels 43; Gaps 10;

OY 6 WDARVAGGRRYRINN--VGAETACIE-VGLETGNTTRADHDNG----- 50
DB 110 WSCGGANQOMRVNSDGTIVGVESGLCEAAGATNGTAVQLMTCGGGNQKWTGLTGT 169
OY 51 ---NNVAAPYATFGCHMGACTNSGLPRVQELSDVRSW-FLTPPTGRMAAADYWF 106
DB 170 PRPDGTICALPSTY---RW-----STGV-----LAQPKSVAALKDFTYTHNGRHLVYG 216
OY 107 SPVTSNGNGYSGAELMIWLNMGVWPGGSRVATVELACA-----TWEEVYADMDW 158
DB 217 S--TSSGSSV-GSMVFSPFTNMSDMASAGNANNOAAVATLTFYFAPKNIVLAWQGSW 273
OY 159 NYIAYRRTPPT 169
DB 274 PTYRTSSDPT 284

RESULT 10
YBPM_ECOLI
ID YBPM_ECOLI STANDARD: PRT; 468 AA.
AC P75733;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ybfm.
GN YBPM OR B0681.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 748-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC DR EMBL: AE000172; AAC73775.1; -
CC DR EMBL: D90707; BAA35329.1; ALT_INIT.
CC DR EcoGene: EG13659; YbFM.
CC KX Hypothetical protein: Complete proteome.
CC SQ SEQUENCE 468 AA; 52780 MW; 934B5A778A1C2E4 CRC64;

Query Match 6.7%; Score 83; DB 1; Length 468;
Best Local Similarity 23.6%; Pred. No. 4.5;
Matches 46; Conservative 25; Mismatches 70; Indels 54; Gaps 10;

OY 44 RADHDGNNVAAPYATFGCHMGACTNSGLPRVQELSDVRSWT----- 89
DB 281 RSYNDYDGTAWLQALTFG-----YRADVYDLRLGKVVAKDGOOGYFLOR 327
OY 90 LTPPTGRMAAADYWFSPVTSNGNGYSGAELMIWLNMGVWPGGSRVATV 142
DB 328 MTP-TYASSNGRLDIMW---DNRSDFNAGEKAVFFGAMYLKNNM---LPFAAGAST 379
OY 143 ELGATWEVYADMDNYIAY--RRTPTTSVSELDLKAETDAVARGYIRPMYLAHVE 200
DB 380 VYA---WDAPKATWQSNPDAYYDKNRTIESAVSLDAVYIQGRAKGTW---FKLHFT 433
OY 201 ---TGFELEMGAG 211
DB 434 YDNHSDIPSWGGYG 448

RESULT 11
GUXB_CELFI
ID GUXB_CELFI STANDARD: PRT; 1090 AA.
AC P50899;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellulohydrolase B)
DE (1,4-beta-cellulohydrolase B) (CBP120).
GN CBHB OR CENE.
OS Cellulomonas fami.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.
RC STRAIN-ATCC 484;
RX MEDLINE=96003898; PubMed=7575482;
RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Cellulohydrolase B, a second exo-cellulohydrolase from the
RT cellulytic bacterium Cellulomonas fami.";
RL Biochem. J. 311:67-74(1995).
RN [2]
RP SEQUENCE OF 54-75.
RX MEDLINE=93209933; PubMed=8458833;


```

DR SMART: SM00450; RHOD: 1.
DR PROSITE; PS00380; RHODANESE_1; PARTIAL.
DR PROSITE; PS00683; RHODANESE_2; 1.
KW Transferase.
FT NON_TER 1
FT ACT_SITE 172 172 BY SIMILARITY.
SQ SEQUENCE 225 AA; 25085 MM; B15A552DFC4BA95B CMC64;

Query Match 6.6%, Score 81.5; DB 1; Length 225;
Best Local Similarity 27.3%; Pred. NO. 2.7;
Matches 44; Conservative 20; Mismatches 56; Indels 41; Gaps 9;

Oy 1 ELGCGMDARDAVAGGARVYINNMWGAETACIEVGEE--TGNF-ITRRADHNGNNVAAYP 57
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 36 ELFGISDVRLLNGR-----DANMAEEDTSTVYPERYSANTPYVERDEQRAFALEV- 89
    - - - - - LGSILTQSGGM-----TLVDVTRPSEFSGL-----DEHGNTSNGVLRG 128

Oy 58 AIFGCHWGACTSNSGLPRVQELSDVTSMTLPTTGRNNAAYDIWFSPVTSNG--N 114
    - - - - - LGSILTQSGGM-----TLVDVTRPSEFSGL-----DEHGNTSNGVLRG 128
Db 90 -----LGSILTQSGGM-----TLVDVTRPSEFSGL-----DEHGNTSNGVLRG 128

Oy 115 GYSGGAEIMVIMNMGWGVMPGGSRATVETLGAATGEWYAD 155
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 129 GHIPGA---INLWSDAVLPNGNFTREL-----DKLYAD 161

RESULT 14
VGL2_CVPPR
ID VGL2_CVPPR STANDARD; PRT: 1447 AA.
AC 002167;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE E2 glycoprotein precursor (Spike glycoprotein) (peplomer protein) .
GN S.
OS Porcine transmissible gastroenteritis coronavirus (strain Pur6-MAD) .
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
ON NCBI_TaxID=35739;
RX 11
RP SEQUENCE FROM N.A.
MEDLINE=92410651; Pubmed=1326823;
RX Sanchez C.M., Gebauer F., Sune C., Mendez A., Dopazo J., Enjuanes L.;
RT "Genetic evolution and tropism of transmissible gastroenteritis
    coronaviruses."
RL Virology 190:92-105(1992) .
CC CC
CC -I- FUNCTION: THE PEPLMER PROTEIN MEDIATES THE BINDING OF VIRIONS
    TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
    AND IN SYNCTYOTIC FORMATION.
CC CC
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL Outstation -
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    CC modified and this statement is not removed. Usage by and for commercial
    CC entities requires a license agreement (See http://www.isb.ch/announce/
    or send an email to license@isb-sib.ch).
CC CC
DR EMBL; M94101; AAA47109.1; .
DR InterPro; IPR002551; Corona_S1.
DR InterPro; IPR002552; Corona_S2.
DR Pfam; PF01600; Corona_S1; 1.
DR Pfam; PF01601; Corona_S2; 1.
KW Glycoprotein; Envelope protein; Transmembrane; Signal.
FT SIGNAL 1 16
FT CHAIN 17 1447 BY SIMILARITY.
FT DOMAIN 17 1388 E2 GLYCOPROTEIN.
FT TRANSMEM 1389 1408 EXTRACELLULAR (POTENTIAL) .
FT DOMAIN 1409 1447 POTENTIAL.
FT DOMAIN 1409 1430 CYTOSOLASMIC (POTENTIAL) .
FT DOMAIN 1405 465 CYS-RICH.
FT CARBOHYD 26 26 RECEPTOR-BINDING (POTENTIAL) .
FT N-LINKED (GLCNAC. . .) (POTENTIAL) .

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:57:39 ; Search time 45.43 Seconds

(without alignments)
845,364 Million cell updates/sec

Title: US-10-003-759-2_COPY_40_261

Perfect score: 1234
Sequence: 1 ELICRMDARDVAGGRVYN.....PELMEGAGLSADFSVTYQ 222

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_RVIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1106.5	89.7	260	2	033897	033897 rhodothermu
2	396.5	32.1	371	2	09KIH1	09KIH1 streptomyc
3	352	28.5	382	2	059963	059963 streptomyc
4	348.5	28.2	377	2	008468	008468 streptomyc
5	345	28.0	381	2	09RXY3	09RXY3 streptomyc
6	343	27.8	381	2	054331	054331 streptomyc
7	333	27.0	384	2	09X602	09X602 streptomyc
8	242.5	19.7	264	2	031030	031030 pectobacter
9	204.5	16.6	244	2	09KXSA	09KXSA streptococ
10	200.5	16.2	274	16	060033	060033 thermotoga
11	196.5	15.9	274	2	P66492	P66492 thermotoga
12	194	15.7	151	16	053438	053438 mycobacteri
13	178.5	14.5	238	3	094218	094218 aspergillus
14	176.5	14.3	234	3	000095	000095 trichoderma
15	172.5	14.0	239	3	013454	013454 aspergillus
16	161.5	13.1	239	3	074705	074705 aspergillus

17	160.5	13.0	319	1	09V2T0	09V2T0 pyrococcus
18	156.5	12.7	263	2	09AN41	09AN41 bradyrhizob
19	153	12.4	257	2	008428	008428 thermotoga
20	153	12.4	257	2	060032	060032 thermotoga
21	153	12.4	257	2	P66491	P66491 thermotoga
22	153	12.4	258	16	09S5X8	09S5X8 thermotoga
23	142.5	11.5	332	17	09TVG7	09TVG7 sulfolobus
24	139	11.3	334	17	09TX08	09TX08 sulfolobus
25	136	11.0	239	3	09P8N6	09P8N6 caldicellul
26	106.5	8.6	361	2	052375	052375 caldicellul
27	98	7.9	608	2	093GM7	093GM7 salmone
28	95.5	7.7	322	17	09TVS7	09TVS7 sulfolobus
29	94.5	7.7	516	3	09URR7	09URR7 penicillium
30	94.5	7.7	516	3	09URR6	09URR6 penicillium
31	94	7.6	112	5	09V6J2	09V6J2 drosophila
32	93.5	7.6	673	16	007178	007178 mycobacteri
33	92.5	7.5	840	10	093X57	093X57 fragaria an
34	91.5	7.4	690	5	09XWC5	09XWC5 caenorhabdi
35	91	7.4	950	16	09A4B3	09A4B3 caulobacter
36	90.5	7.3	516	3	09URR5	09URR5 penicillium
37	89.5	7.3	574	2	086876	086876 streptomyc
38	89.5	7.3	1287	2	093105	093105 burkholderi
39	89	7.2	242	5	096687	096687 lumbricus b
40	88	7.1	928	10	09LTV0	09LTV0 arabidopsis
41	87	7.1	604	17	09HRM4	09HRM4 halobacteri
42	86.5	7.0	360	2	P77853	P77853 dictyoglomu
43	86.5	7.0	457	2	052779	052779 clostridium
44	86.5	7.0	683	2	087119	087119 clostridium
45	86	7.0	842	2	09XCV4	09XCV4 cellulomona

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	260 AA.
033897	033897	033897		
AC	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CELLULOSE (EC 3.2.1.4).			
GN	CELA.			
OS	Rhodothermus marinus (Rhodothermus obamensis).			
OC	Bacteria; CFB group; Rhodothermus group; Rhodothermus.			
OX	NCBI_TaxID=29549;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATN-TT1378;			
RX	MEDLINE=98242392; PubMed=9581291;			
RA	Hallforsdottilir S., Thorolfssdottilir E.T., Spilliaert R., Johansson M.,			
RA	Thorbjarnardottir S.H., Palsdottir A., Hreggvidsson G.O.,			
RA	Kristjansson J.K., Holst O., Eggertsson G.;			
RT	*Cloning, sequencing and overexpression of a Rhodothermus marinus gene			
RT	encoding a thermostable cellulase of glycosyl hydrolase family 12.			
RL	Appl. Microbiol. Biotechnol. 49:277-284 (1998).			
DR	EMBL; U72637; AAB55594.1; -			
DR	InterPro; IPR002594; Glyco_hydro.12.			
DR	Pfam; PF01670; Glyco_hydro.12; 1.			
DR	Prodom; PD004316; Glyco_hydro.12; 1.			
KW	Hydrolase; Glycosidase.			
SQ	SEQUENCE 260 AA; 28770 MW; 94F197DB7D0D247A CRC64;			

Query Match	89.7%; Score 1106.5; DB 2; Length 260;
Best local Similarity	88.6%; Pred. No. 3.3e-87;
Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;	
OY	1 ELICRMDARDVAGGRVYNINNGAETACIEVGETTIRADHDGNNVAAYPAIY 60
DB	40 ELICRMDARDVAGGRVYNINNGAETACIEVGETTIRADHDGNNVAAYPAIY 99
OY	61 FCGHMA-----CTNSGLPRVQELSDVRSWTLPLTTGNNMAAYDWSPVNSGN 114

```
||||| : : | | | : ||||| ||||| |||||
Db 100 FCGHMPARAIRDCARAGAVRRHIELD-----VTPTTGRMNAAYDWFSPVNSGN 152
OY 115 GYSGGALMIWLMNMGVPGSGSRVATVELAGATWEVWADMDWNTIARRTPTTSVSE 174
Db 153 GYSGGALMIWLMNMGVPGSGSRVATVELAGATWEVWADMDWNTIARRTPTTSVSE 212
OY 175 LDKAFIDDAVARGYIRPEWYLHAVETGFELWEGAGLSADESVTVQ 222
Db 213 LDKAFIDDAVARGYIRPEWYLHAVETGFELWEGAGLSADESVTVQ 260

RESULT 2
O9KIH1 PRELIMINARY; PRT; 371 AA.
AC O9KIH1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOSE 12A.
GN
OS Streptomyces sp. 11AG8.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1334352;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=11AG8;
RA van Solingen P., Meijer D., van der Kleij W.A.H., Barnett C.C.,
RA Bolle R., Power S.D., Jones B.E.;
RT "Cloning and expression of an endocellulase gene from a novel
RT Streptomyces isolated from an East African soda lake.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF233376; AAF91283.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR002584; Glyco_hydro_12.
DR InterPro: IPR001230; Prenyltn.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR PROSITE: PS00294; PRENYLATION; UNKNOWN_1.
SO SEQUENCE 371 AA; 38481 MW; 0E1BC428BA148914 CRC64;

Query Match 32.1%; Score 396.5; DB 2; Length 371;
Best Local Similarity 39.7%; Pred. No. 3.9e-26;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

OY 1 ELGGRDADVAGRRYRINNWGAETACIEVGLTGN-FTTRADHDGNNVA--AYP 57
Db 34 QICDRIGTTTQ-DRYVVONNRMGTATOCIN--TGNGFETQADGSPVTKAPKSY 89
OY 58 AIFGCHMGACTNSGLPRVOELSDVTSWTLPTTGRMNAAYDIWFSPVNSGNGYS 117
Db 90 SYVDGCHYGCAPRTTLPRHISISGSAPSVSYRTGNGVYNAAYDIWLDP-TPRNGVN 148
OY 118 GGAELMIWLMNMGVPGSGSRVATVELAGATWEVWADMDWNTIARRTPTTSVSELDL 177
Db 149 -RTEIMWTFNRVGPVQIPISPVGTAVHVGSGSEVMTGNSGNDVISELAPSAISSMSFDV 207
OY 178 KAFIDDAVARGYIRPEWYLHAVETGFELWEGAGLSADESVTV 221
Db 208 KDFVDAVSHGLATPDWYLTISQAGFEPWEGGTGLAVNSFSNAV 251

RESULT 3
O59963 PRELIMINARY; PRT; 382 AA.
AC O59963;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOSE (EC 3.2.1.4).
GN EGLS.
```

```
OS Streptomyces rochei (Streptomyces parvulus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1928;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RX MEDLINE=95011642; PubMed=7523249;
RA Perito B., Hanhart E., Irdani T., Iqbal M., McCarthy A.J.,
RA Mastromel G.;
RT "Characterization and sequence analysis of a Streptomyces rochei A2.";
RL Gene 146:119-124(1994).
DR EMBL: X73953; CAA52139.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF00553; CBD_2; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Hydrolase; Glycosidase.
SO SEQUENCE 382 AA; 39398 MW; 21C014342EFC6565 CRC64;

Query Match 28.5%; Score 352; DB 2; Length 382;
Best Local Similarity 36.2%; Pred. No. 2.6e-22;
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

OY 11 VAGGRYRINNWGAETACIEVGLTGNFTTRADHDGNNVA--AYPIYRGCHMGAC 68
Db 49 VIGGRYVONNRNGTSATOCV-TATDSG-FRYTQADGSPVTKAPKSYSVFNGCHYTC 106
OY 69 TNSGGLPRVOELSDVTSWTLPTTGRMNAAYDIWFSPVNSGNGYSGGALMTLWN 128
Db 107 SPTALPARISGISAPSSISYGFVNAYNASTYDMLDPTPTDG--VNRTEIMWFR 164
OY 129 NGCVPGSGSRVATVELAGATWEVWADMDWNTIARRTPTTSVSELDKAFIDDAVARG 188
Db 165 VGIQIPIGSQVGTASVAGRTWEVWSGNGTNDVLSFVARSAMSWSFDVDFRATVARG 224
OY 189 YIRPEWYLHAVETGFELWEGAGLSADESVTV 221
Db 225 LAGNDWYLTISQAGFEPWONGAGLAVNSFSSTV 257

RESULT 4
O08468 PRELIMINARY; PRT; 377 AA.
AC O08468;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CEL2 (EC 3.2.1.4).
GN CELA2.
OS Streptomyces halstedii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1944;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JMB;
RX MEDLINE=97307849; PubMed=9182697;
RA Garcia-Salas A.L., Fernandez-Abalos J.M., Sanchez P., Ruiz-Arribas A.,
RA Santamaria-Sanchez R.I.;
RT "Two genes encoding an endoglucanase and a cellulose-binding protein
RT are clustered and co-regulated by a TTA codon in Streptomyces
RT halstedii JMB.";
RL Biochem. J. 324:403-411(1997).
DR EMBL: U51222; AAC45429.1; -.
DR HSSP: P07986; 1EXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002584; Glyco_hydro_12.
DR Pfam: PF00553; CBD_2; 1.
```


DR Pfam: PF01670; Glyco_hydro.12; 1.
DR Prodom: PD004316; Glyco_hydro.12; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 377 AA; 38829 MW; 9892191B7A1EC280 CRC64;

Query Match 28.2%; Score 348.5; DB 2; Length 377;
Best Local Similarity 35.9%; Pred. No. 5.2e-22;
Matches 80; Conservative 43; Mismatches 91; Indels 9; Gaps 8;

QY 2 LGRMDADVAGGRYRINNVWGAEETACIEVGLFETGNFTIRADHDNGNVA--AYPAI 59
DB 41 VCGQYSTTIO-GRVYQNNRWGASAPQCV-TATDSG-FRVYQADGAPFNKAPKSTPSV 97
QY 60 YFGCHGACTSNGLPRRVOELSDVTSWTLPTTGRRMAAYDIFSPVTSNGNGYSGC 119
DB 98 FNGCHYTCSPGNLPAQVSGIASAPSSISYGVGSAVYNASDYIMLP-TPKKNGN-R 155
QY 120 AELMILNMGVMPGSRVAVELAGATWEVYADMDN-YIAYRRTPTTSVSELDK 178
DB 156 TEIMILNKVGPITQIGSQAGTASVGRFTQVNRGNSNDVISFVAPSAVASMS-FDVM 214
QY 179 AFIDDAVARGIRPEWTLHAVETGFEFMEGAGLRSDAFSVTV 221
DB 215 DFVRNTIARGMAQNNMYLTISVQGFEPQMGAGLAVNSFSSTV 257

RESULT 5
Q9RJY3 PRELIMINARY; PRT; 381 AA.

AC Q9RJY3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE SECRETED CELLULOSE B.
GN CELB.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Krieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinash H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL EMBL: AL133210; CAB61599.1; -.
DR HSSP: P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR Prodom: PD004316; Glyco_hydro.12; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
SQ SEQUENCE 381 AA; 39199 MW; 97CD8F58679E4EC CRC64;

Query Match 28.0%; Score 345; DB 2; Length 381;
Best Local Similarity 36.5%; Pred. No. 1e-21;
Matches 77; Conservative 31; Mismatches 97; Indels 6; Gaps 4;

QY 14 GRRVYNNWGAETACIEVGLFETGNFTIRADHDNGNVA--AYPAIFGCHGACTSN 71
DB 55 GRRVQNNRWGASTATQCV-TATDTG-FRVYQADGSAPTNAPKSPSVFNGCHYTCSPG 112
QY 72 SGLPRRVOELSDVTSWTLPTTGRRMAAYDIFSPVTSNGNGYSGGAEIMTLWMNG 131
DB 113 TALPVRIDTVSAAPSSISYGFVDAVYNASDYIMLPARTDG--VNGTEIMTFNRVGP 170
QY 132 VMEGSRVATVELAGATWEVYADMDNYYIAYRRTPTTSVSELDKAFIDDAVARGYR 191
DB 171 IOTIGSPVGTASVGRTEWEMSGNSNDVLSFVAPSAISGNSFDVMDVRAIYARGLA 230
QY 192 PEWYLAHETGFEFMEGAGLRSDAFSVTVQ 222
DB 231 NDVYLTISVQGFEPQMGAGLAVNSFSSTVE 261

RESULT 6
Q54331 PRELIMINARY; PRT; 381 AA.

AC Q54331;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOSE B PRECURSOR.
GN CELB.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-66;
RA Wittmann S., Sharek F., Kluepfel D., Morosoli R.;
RT "Purification and characterization of the CelB endoglucanase from
RT Streptomyces lividans 66 and DNA sequence of the encoding gene.";
RL Appl. Environ. Microbiol. 60:1701-1703(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-66;
RA Sharek F.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 004629; AAB71950.1; -.
DR HSSP: P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR Prodom: PD004316; Glyco_hydro.12; 1.
DR PROSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 40 POTENTIAL.
FT CHAIN 41 381 CELLULOSE B.
SQ SEQUENCE 381 AA; 39239 MW; A7E99BF590FA24EC CRC64;

Query Match 27.8%; Score 343; DB 2; Length 381;
Best Local Similarity 36.5%; Pred. No. 1.5e-21;
Matches 77; Conservative 31; Mismatches 97; Indels 6; Gaps 4;

QY 14 GRRVYNNWGAETACIEVGLFETGNFTIRADHDNGNVA--AYPAIFGCHGACTSN 71
DB 55 GRRVQNNRWGASTATQCV-TATDTG-FRVYQADGSAPTNAPKSPSVFNGCHYTCSPG 112
QY 72 SGLPRRVOELSDVTSWTLPTTGRRMAAYDIFSPVTSNGNGYSGGAEIMTLWMNG 131
DB 113 TDLPRVLDTVSAAPSSISYGFVDAVYNASDYIMLPARTDG--VNGTEIMTFNRVGP 170
QY 132 VMEGSRVATVELAGATWEVYADMDNYYIAYRRTPTTSVSELDKAFIDDAVARGYR 191
DB 171 IOTIGSPVGTASVGRTEWEMSGNSNDVLSFVAPSAISGNSFDVMDVRAIYARGLA 230

OY 192 PENYLAHETGFELEMGAGLRADFSVTQ 222
 DB 231 NDWYLTSSVQGFEPQMGAGLAVNSFSSTVE 261

RESULT 7
 OYX602 PRELIMINARY; PRT: 384 AA.

ID OYX602
 AC OYX602;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CELULYASE.
 GN CELS.

OS Streptomyces viridosporus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=67581;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-T7A;

RA Ramachandran S., Crawford D.L.;

RT "Characterization and sequence analysis of two genes involved in
 cellulase degradation in Streptomyces viridosporus T7A, and its
 expression in *Escherichia coli*."

RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF130408; AAD25090.1; -

DR HSSP: P07986; 1EXG.

DR Interpro: IPR001919; CBD_2.

DR Interpro: IPR002594; Glyco_hydro_12.

DR Pfam: PF00553; CBD_2; 1.

DR Pfam: PF01670; Glyco_hydro_12; 1.

DR ProDom: PD004316; Glyco_hydro_12; 1.

SO SEQUENCE 384 AA; 40918 MW; D3968BBE8DEDE65 CRC64;

Query Match 27.0%; Score 333; DB 2; Length 384;
 Best Local Similarity 35.0%; Pred. No. 1.1e-20;
 Matches 77; Conservative 34; Mismatches 93; Indels 16; Gaps 6;

OY 14 GRRVYNNVWGAEETACIEVGLTGNFTTIRADHNGNNVA--AYAIYGCWMGACTSN 71

DB 52 GRVYVNNRNGSSPCV-TATDTG-FRLQADGSVPTNAPKSPSYFVNGCHYTNCSPG 109

OY 72 SGLPRVQELSDVRT-----SWTLFTPTTGR-----WNAAYDIWFSPVTSNGYSGAE 121

DB 110 TKLPATISGSSAPARITXGSSAPSSISYCGAVYNASDIWIDPTPTD--VNRTD 167

OY 122 LMIWLNMGVMPGSRVATVELAGATWEVYVADWDMNYIAYRRTPPTTSVSELDLKAFT 181

DB 168 IMIMFKVKGPIQIGSQVGTATVCGRTQWVSGNGSNDLFSFVAPSAIESMFDVDFV 227

OY 182 DDAVARGYIRPEWYLAHETGFELEMGAGLRADFSVTQ 221

DB 228 RETVARGMAQNDWYLTSSVQGFEPQMGAGLAVNSFSSTV 267

RESULT 8
 OYX602 PRELIMINARY; PRT: 264 AA.

ID OYX602
 AC OYX602;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
 GN CELB.

OS Pectobacterium carotovorum subsp. carotovorum.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Pectobacterium.

NCBI_TaxID=555;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-LY34;
 RA MEDLINE=98096373; PubMed=9434760;
 RX Park Y.W., Lim S.T., Cho S.J., Yun H.D.;

RT "Characterization of *Erwinia carotovora* subsp. *carotovora* LY34 endo-

RT 1,4-beta-glucanase genes and rapid identification of their gene

RT products."

RT Biochem. Biophys. Res. Commun. 241:636-641(1997).

DR EMBL: AF025769; AAC02965.1; -

DR Interpro: IPR002594; Glyco_hydro_12.

DR Pfam: PF01670; Glyco_hydro_12; 1.

DR ProDom: PD004316; Glyco_hydro_12; 1.

DR Signal; Hydrolase.

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36

FT CHAIN 37 264

FT SIGNAL 1 36


```

Db 44 WNAKSYEG-----ETWLKDEGEKVOEYADIVNYLONPD-----SWVHGYPETLYGKXP 92
QY 65 WCACTSNSSG---LPRVQELSD---VRTSWTL-----TPITGRMNAAYDIWESPTVSGN 114
Db 93 WAA--HNSGTEILPVKVKDLPDFYVLDYISIVWENDLPT-----NLAMEWITRKPPOTS 145
QY 115 GYSGAGELMIWLMWNGVMDGSGRV-----AVVELAGA---TWVYADMDMYIAVART 166
Db 146 VSSGDEIVWVF--YNNILMPGSGKQDEFTTIEINSPTVEKMDVYFAPMGWIDYLAFLRT 204
QY 167 TPTTYSV--ELDLKAFIDDA---VARGIIRPEWYLHAVETGFELMEGA-----210
Db 205 TPKKDGVRKFNVDPEVKAELVIKHSITRAVE---NPEDEMYCWEIGETEGEPNTTAAKF 261
QY 211 GLRSADFSVTV 221
Db 262 GWTFKDFSEVI 272

```

```

RESULT 12
053438 PRELIMINARY; PRT; 151 AA.
AC 053438;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE HYPOThETICAL 16.4 KDA PROTEIN.
GN RV1090 OR MTV017.43.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Horsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998)
DR EMBL: AL021897; CAI17206.1; -
DR TuberCulist: RV1090; -
DR TuberCulist: IPR002594; Glyco_hydro.12.
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR ProDom: PD004316; Glyco_hydro.12; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 151 AA; 16404 MW; 8BCC10CF9C98E54 CRC64;

```

```

Query Match 15.7%; Score 194; DB 16; Length 151;
Best local Similarity 33.3%; Pred. No. 3e-09;
Matches 51; Conservative 16; Mismatches 78; Indels 8; Gaps 4;
QY 72 SGLPRVQELSDVRTSWTLPTITGRMNAAYDIW--SPVNSGNGYSGAGELMIWLNWN 129
Db 3 TNPTEVGQILSAPTSIDYNYPTTGVWDASXDICLSTPTKTCVN---QOELIMENHQ 58
QY 130 GGVMPGSGRNATVELAGATMEVYADMDW--NTAYARRTPPTTSVELDLKAFIDDAVARG 168
Db 59 GSTQPGVSGPNTTIEGKNEVMDGSGNMMNAAYATEP--IEWSFSDVFDHTATME 117
QY 189 YIRPEWYLHAVETGFELMEGAGLRSADFSVTV 221
Db 118 PITDSWTLTISIRAGLEPMSDGLVDSFSRAKY 150

```

```

RESULT 13
094218 PRELIMINARY; PRT; 238 AA.
ID 094218;
AC 094218;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-GLUCANASE PRECURSOR.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KSM 510;
RC MEDLINE=99102417; PubMed=9884411;
RA Pauly M., Andersen L.N., Kauppinen S., Kofoed L.V., York W.S.,
RA Albersheim P., Darvill A.;
RA "A xyloglucan-specific endo-beta-1,4-glucanase from Aspergillus
RT aculeatus: expression cloning in yeast, purification and
RT characterization of the recombinant enzyme."
RL Glycobiology 9:93-100(1999).
DR EMBL: AF043595; AAD02275.1; -
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR ProDom: PD004316; Glyco_hydro.12; 1.
KW Signal.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 238 XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-
FT FT GLUCANASE.
SQ SEQUENCE 238 AA; 25158 MW; FCCA6746D9AEC1B1 CRC64;

```

```

Query Match 14.5%; Score 178.5; DB 3; Length 238;
Best local Similarity 25.4%; Pred. No. 1.1e-07;
Matches 63; Conservative 39; Mismatches 89; Indels 57; Gaps 12;
QY 1 ELGRMDARVAGGRVYINNVWGAET--ACCEVGLTGNFTIRADHGNNAVAP 57
Db 22 DFCGQMDT--ATAGDPTLYNDLWGESAGTGSQC-----TGVDSYSGDTIA--- 64
QY 58 AITFGCHWCACTNSGLPRVQELSDVRTSWTLPTI-----TTGRN-----N 99
Db 65 -----WHTSMWSGSSSSSVK--SYVNALITFTPOLNCISSIPTTWKSYSGSSIVAD 115
QY 100 AAYDIWESPTVNSGNGYSGAGELMIWLNNGVMP---GSSRVATVELAGATMEVYA-D 155
Db 116 VAYDTFLAETASGSSKY---ETVWNLALGGAGPISSTGSTITATPTTGVNKKLSGPN 171
QY 156 WDNNTYARRTPPTTSVELDLKAFIDDAVARGYIRPEWYLHAVETGFELMEG--GAGLRS 214
Db 172 GDTTVVSFVADSTTESFSG--DLNDFTYLVNDCGVDELITLTLEAGTEPFGSNKKLV 230
QY 215 ADFSVTV 222
Db 231 SEYSISIE 238
RESULT 14
000095 PRELIMINARY; PRT; 234 AA.
ID 000095;
AC 000095;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)
DE ENDO-BETA-1,4-GLUCANASE (EC 3.2.1.4).
GN EGL.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 12:44:21 ; Search time 6219.51 Seconds
(without alignments)
2644.623 Million cell updates/sec

Title: US-10-003-759-3

Perfect score: 786
Sequence: 1 atgaacgcatcgcgtgcggt.....ttccgtaacggtgcagtag 786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb_da: 2: gb_hlg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: gb_da: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vl: 30: em_hlg_hum: 31: em_hlg_inv: 32: em_hlg_other: 33: em_hlgo_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	786	100.0	786	6	AX339682	AX339682 Sequence
2	786	100.0	1877	6	AX339680	AX339680 Sequence
3	746.8	95.0	1874	6	RK072637	U72637 Rhodothermu
4	131.8	16.8	1116	6	AF233376	AF233376 Streptomy
5	131.8	16.8	1116	6	AR129926	AR129926 Sequence
6	131.8	16.8	1116	6	AR168360	AR168360 Sequence
7	105.8	13.5	1716	6	AR168362	AR168362 Sequence
8	97.8	12.4	3527	1	SH051222	U51222 Streptomyce
9	96.2	12.2	1470	1	SREGLS	X73953 S.rochei eg
10	89	11.3	1311	1	AF130408	AF130408 Streptomy
11	87.6	11.1	1565	1	SLU04629	U04629 Streptomyce
12	87.6	11.1	41782	1	SCG11A	AL13220 Streptomy
13	69	8.8	11202	1	AE005123	AE005123 Halobacte
14	68.8	8.8	4262	1	AF335723	AF335723 Burkholde
15	64.4	8.2	14483	1	AE006992	AE006992 Mycobacte
16	64.4	8.2	67200	1	MTV017	AL021897 Mycobacte
17	57.6	7.3	1983	6	AX047403	AX047403 Sequence
18	55	7.0	36368	1	SC9B5	AL035206 Streptomy
19	54.8	7.0	346897	1	AP002995	AP002995 Mesorhizo
20	54.6	6.9	45313	1	SCD95A	AL357432 Streptomy
21	54.6	6.9	125020	9	AF429315	AF429315 Homo sapi
22	54.4	6.9	38995	1	SCF34	AL109974 Streptomy
23	54	6.9	34182	1	SC111	AL096849 Streptomy
24	53.6	6.8	77534	1	AF235504	AF235504 Streptomy
25	53	6.7	33779	1	SCG03	AL096882 Streptomy
26	52.8	6.7	125020	9	AF429315	AF429315 Homo sapi
27	52.2	6.6	1849	1	SGSTRB1F	X78973 S.galbus DS
28	52	6.6	349116	1	AP003003	AP003003 Mesorhizo
29	51.4	6.5	3644	5	AF034576	AF034576 Gallus ga
30	51.2	6.5	22115	1	SCC22	AL096839 Streptomy
31	50.6	6.4	5525	1	SGSTRB1FG	X78974 S.glaucisce
32	50.6	6.4	14596	1	AE007168	AE007168 Mycobacte
33	50.6	6.4	25459	1	SGAJ6985	AJ06985 Streptomy
34	50.6	6.4	30302	1	MTCV6G11	292774 Mycobacteri
35	50.6	6.4	40806	1	MSGV23	AD000016 Mycobacte
36	50	6.4	29625	1	SC422A	AL159178 Streptomy
37	50	6.4	36394	1	SCD35	AL160312 Streptomy
38	50	6.4	40104	1	SCF11	AL12662 Streptomy
39	49.8	6.3	36734	1	SC6610	AL049497 Streptomy
40	49.6	6.3	26555	1	SC261	AL391014 Streptomy
41	49.6	6.3	146555	2	OSJN00013	AL606448 Oryza sat
42	49.6	6.3	215050	1	AL646057	AL646057 Ralstonia
43	49.4	6.3	150814	2	AP003541	AP003541 Oryza sat
44	49.2	6.3	12130	1	AE005133	AE005133 Halobacte
45	49.2	6.3	41173	1	SC141	AL132648 Streptomy

ALIGNMENTS

RESULT	1	786 bp	DNA	linear	PAT 10-JAN-2002
AX339682	AX339682	Sequence 3 from Patent WO0196382.			
LOCUS	AX339682				
DEFINITION	AX339682				
ACCESSION	AX339682				
VERSION	AX339682.1	GI:18135684			
KEYWORDS					
SOURCE					
ORGANISM	Rhodothermus marinus.				
REFERENCE	1 (sites)				
AUTHORS	Wichner,K.B., Holst,O.P., Hachem,M.Y., Karlsson,E.M. and Hreggvidsson,G.O.				
TITLE	Thermotable cellulase				
JOURNAL	Patent: WO 0196382-A 3 20-DEC-2001;				
FEATURES	Prokaria enf. (IS)				
Source	location/Qualifiers				
	1..786				
	/organism="Rhodothermus marinus"				
	/db_xref="taxon:29549"				

BASE COUNT	141 a	236 c	279 g	130 t
ORIGIN				

Db 1370 GAGGAGCGGTCCTCCCGCGCTACATCCGACCGAGTGTATCGATGCGGTGAGACG 1429
OY 721 ggcttgaactctggagagcgagccgctctgcaagcgacgatttccgtaacggtg 780
Db 1430 GCGTTGCAACTCTGGAGGCGGCGCGGTCTCGAAGCGCCGATTTTTCGTAACGCTG 1489
OY 781 cagtag 786
Db 1490 CAGTAG 1495
RESULT 3
LOCUS RMU72637 1874 bp DNA linear BCT 11-MAY-1999
DEFINITION Rhodothermus marinus cellulase (celA) gene, complete cds.
ACCESSION U72637
VERSION U72637.1 GI:2304960
KEYWORDS Rhodothermus marinus.
SOURCE Rhodothermus marinus.
ORGANISM Bacteria; CF8 group; Rhodothermus group; Rhodothermus.
REFERENCE 1 (bases 1 to 1874)
AUTHORS Halldorsdottir, S., Thorolfsson, E.T., Spilliaert, R.,
Johansson, M., Thordarson, S.H., Palsdottir, A.,
Hrengsson, G.O., Kristjansson, J.K., Holst, O. and Eggertsson, G.
Cloning, sequencing and overexpression of a Rhodothermus marinus
gene encoding a thermostable cellulase of glycosyl hydrolase family
12
JOURNAL Appl. Microbiol. Biotechnol. 49 (3), 277-284 (1998)
MEDLINE 98242392
REFERENCE 2 (bases 1 to 1874)
AUTHORS Halldorsdottir, S.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1996) Department of Molecular Biology,
Institution of Biology, Grensasvegur 12, Reykjavik 108, Iceland
FEATURES
source
1. 1874
/organism="Rhodothermus marinus"
/strain="IT1378"
/db_xref="taxon:29549"
710. 1493
/gene="celA"
710. 1492
/gene="celA"
/EC_number="3.2.1.4"
/note="belongs to family H of cellulase catalytic domains;
endo-1,4-beta-glucanase"
/codon_start=1
/product="cellulase"
/protein_id="AAB65594.1"
/db_xref="GI:2304961"
/translation="MNMVRAVIVLSILLFGDMLEPPDNGKEPPEPEPEPEYELGCR
HNPARYAGRRYVNNVKAERNAEFLNGLFENFTITRADHDNKNVAAVPAITFGC
HNAAPAIKDCARAGAVRAHEDLPTPTTRMNAAYDIWSPVNSNGSGAEI
MIMWMNGVMGSGSRVATVELAGATVEVYADMDMNYLAVRTPTTSSEIDLKAF
IDDAVARGYIREMYHAYETGFEIWMGAGARTADESVTVQ"
BASE COUNT 331 a 580 c 631 g 332 t
ORIGIN
Query Match 95.0%; Score 746.8; DB 1: Length 1874;
Best Local Similarity 99.4%; Pred. No. 2.5e-99;
Matches 781; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

OY 121 ctgtgcgacgctggagacgcgcgcga tctggccggggggcgctaccgggtgatacaaac 180
Db 830 CTTGTGGAGACGCTGGGACGCGCGCATGTGCGGGGGGCGCTACGCGGATCAACAAAC 889
OY 181 gtagggcgcgagagacgcgcagtgatgaagtcggaactgtaaacgggcaacttcagc 240
Db 890 GTATGGGGCGCGGAGACCGCCAGTGCATTGAGTGCAGTGCAGAAACGGGGCAACTTCACG 949
OY 241 atcacagggcgagatcacgacaaacgaacgctggccgcatccatccattc 300
Db 950 ATCACAGGGCGCATCAACGACAAACGACAAACGTCGCGCTATCCGGCATCTACCTTC 1009
OY 301 ggggtccactggggcgccctgcacagcaatctcggaattcccgcgcgctgcagagctg 360
Db 1010 GGGTGGCACT -GGCGCCCTGACAGACGAATTCGGGATTG -CGCGGGCGGTGCGAGAGCTG 1067
OY 361 tccgacgtgcgacagacgctgagcgtcaacgcgaatacagaagggcgctggaatgcgcc 420
Db 1068 TCCGAGGTGCGCACGAGCTGAGAG -TCACGCCGATCACGACGGGCGCTGGAATGCCGCC 1126
OY 421 taagacatctggttcagctccgctcacgaattccggaacggctacagcgcgcgcgag 480
Db 1127 TACGACATCTGTGTTCACTCCGCTACGCAATTCGCGACAGCTACAGCGCGCGCGAG 1186
OY 481 ctgatgactgctgactgagaaacggcgcgctgatactgcggcgcgacgcgctgacac 540
Db 1187 CTGATGATCTGGCTGAACCTGGAACGGCGGGGTGATGCGGGCGGACGCCGCGTGGCCAC 1246
OY 541 gtggaactgycggggcgccactggaagctgtgatarccgactcgggaactggaatcac 600
Db 1247 GTGGAACCTGCGCGGGCGCACCTTGGAAGTCTGTATGCCACAGGAGCTGGAATTCATC 1306
OY 601 gctcacggcgacagagccacacacgctgtagcgaagctggaactggaagcctcac 660
Db 1307 GCTTACCGGCGCGACGACGCCACACAGCTGAGTGAAGAGTGAAGCTGAAGGCTTATC 1366
OY 661 gacgacgctgcgcgcgcgctacatccgagcggaagtgtatctgacatgcgtgtgagag 720
Db 1367 GACGACGCGGTCCCGCGCGCTACATCCGCGCGAGTGTGATGCTGATGCGGTGAGAG 1426
OY 721 ggcttgaactctggagagcgagccgctctgcaagcgacgatttccgtaacggtg 780
Db 1427 GCGTTGCAACTCTGGAGGCGGCGCGGTCTCGAAGCGCCGATTTTTCGTAACGCTG 1486
OY 781 cagtag 786
Db 1487 CAGTAG 1492
RESULT 4
LOCUS AF233376 1116 bp DNA linear BCT 02-AUG-2000
DEFINITION Streptomyces sp. 11A68 cellulase 12A (cel12A) gene, complete cds.
ACCESSION AF233376
VERSION AF233376.1 GI:9651812
KEYWORDS Streptomyces sp. 11A68.
SOURCE Streptomyces sp. 11A68.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 1116)
AUTHORS van Solingen, P., Meljer, D., van der Kleij, W.A.H., Barnett, C.C.,
Bolte, R., Power, S.D. and Jones, B.E.
Cloning and expression of an endocellulase gene from a novel
Streptomyces isolated from an East African soda lake
Unpublished
JOURNAL 2 (bases 1 to 1116)
REFERENCE van Solingen, P., Meljer, D., van der Kleij, W.A.H., Barnett, C.C.,
AUTHORS Bolle, R., Power, S.D. and Jones, B.E.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Microbial & Molecular Screening, Genencor
International B.V., Archimedesweg 30, 2333 CN Leiden, The

```
FEATURES
    Netherlands
        Location/Qualifiers
            source      1..1116
                        /organism="Streptomyces sp. 11A68"
                        /strain="11A68"
                        /db_xref="taxon:133452"
            gene        1..1116
                        /gene="cell12A"
            misc_feature 1..93
                        /gene="cell12A"
                        /note="Region: signal sequence"
            CDS          1..1116
                        /gene="cell12A"
                        /note="endocellulase"
                        /codon_start=1
                        /transl_table=11
                        /product="cellulase 12A"
                        /protein_id="AAP91283.1"
                        /db_xref="GI:9651813"
                        /translation="MNSHRSATMTVLVVLASGLALFNAAPDAQANOQICDRYGTITI
                        ODHYVONNRNCTSATOCINVTNGSEITQAGSPNTNCPAPKSPSYDGHYGNCA
                        PRTLLPMRISISGASVSRYTNGVYNAADIMLDPRTNGVARTIMIPNNG
                        PVQPGSPVGTAGHVGSRWETWGSNDVLSPLAPASISSMSPKDEVOAASHG
                        LATPMWLTSLIAGREPMRGGLAVNSPSSAVNMGCGMGTGTPAPACOVSYRHTY
                        PCGFTVDTITNTGTSTPVDGMLDPTLPAGHTVTSYMNALISPASGAVTARSTGSNGR
                        IANGTQSFQGTSSGAGTAPAPAGARLNGTSCIVR"
            misc_feature 94..756
                        /gene="cell12A"
                        /note="Region: catalytic domain"
            misc_feature 757..792
                        /gene="cell12A"
                        /note="Region: linker region"
            misc_feature 793..1113
                        /gene="cell12A"
                        /note="Region: cellulose binding domain"
BASE COUNT    191 a      432 c      340 g      153 t
ORIGIN
Query Match      16.8%; Score 131.8; DB 1; Length 1116;
Best Local Similarity 50.7%; Pred. No. 3.4e-10;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

98  ctgaagcccgagccgacgctgagctgtgagcgtggaagcgcgcgagctggtgcccggg 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77  CGGCTCCCGCCAGGCAACCGACAGATCTGCGACCGCTACGGCACCACCGATCCAGG 136

158  ggcgtacacgggtgatacaacaacgatactggcgcgaggaacggccagtgcatggaagtcg 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137  ACCGGTACGtGtGCAACAAACCGCTGGGGCACACGCCACCGACCATGATCAATGTGA 196

218  gactggaagcggaacttcaacatcacacgagcgcgatcacgaacaacggaacagctgg 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197  CCGGCAACGGTTTCGAGATCAACCCAGGCCGACGTTCCGTCGCGACCAACGGGCCCCGA 256

278  ccgcctatccggcacttacttctgggtgacactggggcgcgctgcaagagaattcggat 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257  AGTCTATTCCTCGTCTAGACGGCTGCCACTACGGCAACTCGCGCCCCCGACGACGC 316

338  tgcgcgagcgctgtcaggaactgtccgaagctgcgaacgaactggaagcgtcaagccgata 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
317  TGCCCATGCGGATCATGCTGATCGGACGGCGCCGACGACATGTCTCTACCGCTACACCG 376

398  cgaagggcgctggaattgcgcctacgacatctgtgtcagtcctccgtcaagaattccgga 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377  GCAACGGCGCTCTACAAACGCGCGTACGACATCTGG-----CTGGACCGACACCCCGCA 430

458  acggctacagcgcgcgcgagctgatactcgtgctgaacttgaagcgcgcgctgatacg 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431  CCAACGGGGTGAACCGGACCGAGATCATGTCTGTTCAACCGGGTCCGCCGCTCCAGC 490

518  cgggcgagcgcgctgtgacacgctggaacttgcggggcgcaacttgggaagctgtgatag 577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db      491  CCATCGGTTGCCCGGTGGCGACGGCCCACTGCGCGCCGACAGCTGGAGGTGTGACCG 550
Oy      578  ccgactggagctggaattacatcgcctaccgagcgcaagcagccaccacagctcgtagcg 637
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      551  GCAGCAACGGTTGCAACGACGATGATCTCTTCTGGCGCCCTCCGGATCAGACAGCTGGA 610
Oy      638  agctggacctgaagccttcacatcagacgagcggtgcgcgagtgatacatccgagcgagtc 697
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      611  GCTTCGACGCTCAAGGACTTCGTGACCAAGCCGCTACGCCACGCGCTGGCCACCCGACT 670
Oy      698  gatatctgatacggtggaagcgagcttcgaactctgggaagcgcgagcggtcttcgga 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      671  GGTACCTCACACACATCCAGCGGCGCTTCGACCGCTTGAGAGGGCGGACCGGTCTGCGCG 730
Oy      758  ggcgcgattttccgtaacgtagca 782
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      731  TGAACCTGTTCTCTCCGCGGTGAA 755

RESULT      5
ARI29926    ARI29926    1116 bp    DNA    linear    PAT 16-MAY-2001
LOCUS       Sequence 2 from patent US 6187577.
DEFINITION  ARI29926
ACCESSION   ARI29926
VERSION     ARI29926.1 GI:14117823
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1116)
AUTHORS    Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,M.
TITLE      Cellulase producing Actinomyces cellulase produced therefrom and
            method of producing same
JOURNAL     Patent: US 6187577-A 2 13-FEB-2001;
            Location/Qualifiers
            source      1..1116
            /organism="unknown"
BASE COUNT    191 a      432 c      340 g      153 t
ORIGIN
Query Match      16.8%; Score 131.8; DB 6; Length 1116;
Best Local Similarity 50.7%; Pred. No. 3.4e-10;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

98  ctgaagcccgagccgacgctgagctgtgagcgtggaagcgcgcgagctggtgcccggg 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
77  CGGCTCCCGCCAGGCAACCGACAGATCTGCGACCGCTACGGCACCACCGATCCAGG 136

158  ggcgtacacgggtgatacaacaacgatactggggcgcgaggaacggccagtgcatggaagtcg 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137  ACCGGTACGtGtGCAACAAACCGCTGGGGCACACGCCACCGACCATGATCAATGTGA 196

218  gactggaagcggaacttcaacatcacacgagcgcgatcacgaacaacggaacagctgg 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197  CCGGCAACGGTTTCGAGATCAACCCAGGCCGACGTTCCGTCGCGACCAACGGGCCCCGA 256

278  ccgcctatccggcacttacttctgggtgacactggggcgcgctgcaagagaattcggat 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257  AGTCTATTCCTCGTCTAGACGGCTGCCACTACGGCAACTCGCGCCCCCGACGACGC 316

338  tgcgcgagcgctgtcaggaactgtccgaagctgcgaacgaactggaagcgtcaagccgata 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
317  TGCCCATGCGGATCATGCTGATCGGACGGCGCCGACGACATGTCTCTACCGCTACACCG 376

398  cgaagggcgctggaattgcgcctacgacatctgtgtcagtcctccgtcaagaattccgga 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
377  GCAACGGCGCTCTACAAACGCGCGTACGACATCTGG-----CTGGACCGACACCCCGCA 430

458  acggctacagcgcgcgcgagctgatactcgtgctgaacttgaagcgcgcgctgatacg 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431  CCAACGGGGTGAACCGGACCGAGATCATGTCTGTTCAACCGGGTCCGCCGCTCCAGC 490

Db      491  CCAACGGGGTGAACCGGACCGAGATCATGTCTGTTCAACCGGGTCCGCCGCTCCAGC 490
```

QY 518 cgggagcgacgagcgctgagccacgtggaactgagccggggccacgtggaagtctggtatg 577
 Db 491 CCATCGCTTCGCCGTCGCGCCGACGTCGCGCCGACCTGGAGTGTGGACCG 550
 QY 578 cgcactgagactggaattacatcgctaccgagcgcaagcccaacgctggtgagcg 637
 Db 551 GCAGCAACGGTTGGAACGACGATCTCTTCTGCGGCCCTCCGCAATACGACGCTGGA 610
 QY 638 agctgagactggaagccttcacagacgaggtgagccgagctacatccggcgaggt 697
 Db 611 GCTTCGACGTCAGAGACTTGTGTCAGACGCGCTGACACGCGCTGCGCACCCGGACT 670
 QY 698 ggtatctgactgagcgagcgagcggtctggaactctggagggcgagggcggtctcgaa 757
 Db 671 GGTACTCTACACCAATCCAGCGCGGCTTCGAGCGGTGGAGGGCGGCGGTGTGGCCG 730
 QY 758 ggcgcgaatttccgtaacggtgca 782
 Db 731 TGAACCTGTTCTCTCCGCGGTGAA 755

RESULT 6
 LOCUS ARI68360 1116 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 2 from patent US 6287839.
 ACCESSION ARI68360
 VERSION ARI68360.1 GI:17904235
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1116)
 AUTHORS Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.
 TITLE Cellulase producing actinomycetes, cellulase produced therefrom and
 method of producing same
 JOURNAL Patent: US 6287839-A 2 11-SEP-2001.
 FEATURES
 source location/Qualifiers
 1..1116
 /organism="unknown"
 BASE COUNT 191 a 432 c 340 g 153 t
 ORIGIN

Query Match 16.8%; Score 131.8; DB 6; Length 1116;
 Best Local Similarity 50.7%; Pred. No. 3.4e-10;
 Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;
 QY 98 ctgagcccgagcgagcgctgagctgtggaagcgctggagcgcgcgatgtgagcggg 157
 Db 77 CGGCTCCCGCCGAGCGACACGAGATCTGCGACCGCTACGGCGACCAACGATCCAGG 136
 QY 158 ggcgtacacgggtgatacaacgatacgatggcgagcgagcgccagtgatgaagtcg 217
 Db 137 ACGGGTACGTTGTCGACGACGCGGTGGGCAACGCGGACACCATGATCATATGTA 196
 QY 218 gactggaacggcgacactcaatcaatcacgagcgagtcacgaacgaacgaacgltg 277
 Db 197 CCGGCAACGGTTTCGAGATCACCGACCGACGTTGCTGTCGACCAACGAGCGCCGA 256
 QY 278 cgcgcataccggcactatcttggtggtgacatggggcgagcgctgacaggaattcgagat 337
 Db 257 AGTCTATCTCCCTGCTGTCAGAGGCGTCCACGTAAGGCACTGCGGCCCGCGACGAGC 316
 QY 338 tgcgcgagcgagctgacgagactgtccgagcg tgcgacgagactgagacgactcaagcgatca 397
 Db 317 TGCCTACGCGGATTCAGTGCATGGGAGCGCGGCCGACCATGTTCTCTACCGCTACACCG 376
 QY 398 cgcgcgagcgagctggaatgagcgactgaacatctggttcaagtcacgactcaaatccggca 457
 Db 377 GCACAGCGCGCTCTCAACGCGCGCTACGACATCTGG-----CTGAGACCCGACACCCCGCA 430
 QY 458 agggctaacagcgcgcgagcgagctgagctggtgaactggaacggggcggtgagtcg 517

Db 431 CCACGCGGTGGAACCGGACCGAGATCATGATCTGTTCAACCGGGTCCGCCGCTCAGC 480
 QY 518 cgggagcgacgagcgctgagccacgtggaactgagccggggccacgtggaagtctggtatg 577
 Db 491 CCATCGCTTCGCCGTCGCGCCGACGTCGCGCCGACCTGGAGTGTGGACCG 550
 QY 578 cgcactgagactggaattacatcgctaccgagcgcaagcccaacgctggtgagcg 637
 Db 551 GCAGCAACGGTTGGAACGACGATCTCTTCTGCGGCCCTCCGCAATACGACGCTGGA 610
 QY 638 agctgagactggaagccttcacagacgaggtgagccgagctacatccggcgaggt 697
 Db 611 GCTTCGACGTCAGAGACTTGTGTCAGACGCGCTGACACGCGCTGCGCACCCGGACT 670
 QY 698 ggtatctgactgagcgagcgagcggtctggaactctggagggcgagggcggtctcgaa 757
 Db 671 GGTACTCTACACCAATCCAGCGCGGCTTCGAGCGGTGGAGGGCGGCGGTGTGGCCG 730
 QY 758 ggcgcgaatttccgtaacggtgca 782
 Db 731 TGAACCTGTTCTCTCCGCGGTGAA 755

RESULT 7
 LOCUS ARI68362 1716 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 4 from patent US 6287839.
 ACCESSION ARI68362
 VERSION ARI68362.1 GI:17904238
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1716)
 AUTHORS Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.
 TITLE Cellulase producing actinomycetes, cellulase produced therefrom and
 method of producing same
 JOURNAL Patent: US 6287839-A 4 11-SEP-2001.
 FEATURES
 source location/Qualifiers
 1..1716
 /organism="unknown"
 BASE COUNT 307 a 611 c 559 g 239 t
 ORIGIN

Query Match 13.5%; Score 105.8; DB 6; Length 1716;
 Best Local Similarity 53.5%; Pred. No. 1.7e-06;
 Matches 269; Conservative 0; Mismatches 227; Indels 7; Gaps 2;
 QY 281 cctatccggcactactctggtggtccactgagggcgctgacaggaattcggtatgc 340
 Db 711 CMTATCCCTCGGTCTACGACGGGTGCGCATACGCAACTCGCGGCCCGACGACGCTGC 770
 QY 341 cgcgcgagcgagcgaggtctgagcgctgagcgagcgagctggaagctgaacgcgatacga 400
 Db 771 CCATGCGGATTCACCTCATGTCGAGCGCGGCCGACGATGTTCTCTACCGCTACACCGCA 830
 QY 401 cgggagcgctggaatgagcgctacgacatctggttcagtcgcgtcaacgaattccggcaag 460
 Db 831 ACGGCGCTTCAACGCGCGCTAGACATCTGCTGCTG-----GACCCGACACCCGCGCA 884
 QY 461 gctacagcgagcgagcgagactgagatctgctgactggaactggaacggcgagctgagtcg 520
 Db 885 ACGGGGTGAACCGGACCGAGATCATGATGTTCAACCGGGTCCGCGCGCTCAGCGCA 944
 QY 521 ggcgcagcgagcgagcgagcgagctggaatgagcgagcgagcgagctggaatgagtcg 580
 Db 945 TCGGTTCCGCGGTGCGGACGCGCCACGCTGCGCGCGCCGACAGTGGAGTGGACCGCA 1004
 QY 581 actgagactggaattacatcgctaccgagcgagcgagcgagcgagctgggtgagcgagc 640
 Db 1005 GCACAGGTTGGAACGACGATATCTCTTCTGCGGCCCTCCGCGATCAGACGCTGAGCT 1064

[illegible]

BASE COUNT 623 a 1270 c 1106 g 528 t
ORIGIN

Query Match	Score	DB	Length
12.48	97.8	1	3527

```
Matches 304; Conservative 0; Mismatches 317; Indels 6; Gaps 1.
```

156 ggggcgctaccggtgatcaacaacgtatggcgcgagaccgccagtcgcatltaggt 2

b 747 GGGCCGCTATGTGTCACAGAACCGCTGGGGCGCGAGTGCCCCCAAGTGCGTCACCGC 8

216 cggaactggaacgagcaacttcacgatcacacgagccgatccagacaaacgagcgaacacgt 2

b 807 TACCGACAGCGGCTTCGGGTGACCCAGGCCGACGGCGCGTCCCCACGAACGGTGCCCC 8

276 ggccgctatcgcgcatactcgggtgcgaactggcgccctgcacgagcaattcggg 3

b 867 GAAGTCTATCCCTCGGTCTTCACGGCTGTCACTACACCAACTGCTCGCCGGGACGAA 9

336 atgccgcgcgtgcagagctgtccgacgtgcgcacgagctggacgctcacgcccgtat 3

b 927 CCTCCGGCGAGTGAGCGGTATCGCCAGCGCCCCCAGCAGCATCTCCTACGGCTTGT 9

396 caccgacggccgctggaatgccgacctacgaacatctgtgtcagttcccgctcaggaattccgg 4

b 987 CGGCAGCGCCGTGTACACGCGTCTGGCTGGACCCCAACCCAGAGAA 1

Y 456 caacgctacagcgcggccgagtgtatcttggcttgaaactggaacgcgcgcgtgat 5

b 1047 CCGGGTGAACC-----GTACCGAGATCATGATCTGGCTCAACAAGTCCGCCCGATCCA 1

516 gccgagcagccgcgtggccaaccgltgaaactggcccgggccaacctgggaagtcctgtta >

b 1101 G C C C A T C G G C T C G C A G G C C G G C A C C G C C T C C C G T C G G C G G G C G C A C C T T G G C A G G T G T G A G 1

576 **tcgcgactg**gacttggaattacatcgctaacggcgcacgacgcccacacgctcggctgag

1161 GGGCAGCAACGGCTCCACGACGTCATCTCCCTTCGTCCGCCCTTCGGCCGTCGCGAGCTG

636 cyagctggaacctgaaagccctcatcgacgaaacggtcgcacccgcgcctacatccgcgcgga b

b 1221 GAGCTTCGACGTCATGGACTTCGTCGGGAACACCATCGCCCCGGCATGGCCGAGAACAA 1

696 gtggtatctcgtcgttgagacggtcttcgaaactctggagggcggggccggcctcgcg /

1281 CTGGTATCTCACCAGTGTCCAGGCAGGATTCGAGCCGTCGACGAACGGTGCCTGGACATCGC

Y 756 aagcgccgatttttcgcgtaacggtgca 782

D 1341 GGTGAACCTCCTTCCTCCTCAACCGTGA 1367

RESULT 9

REGLS	OCIS	SPECTS
1470 bp	DNA	11near BCT 16-D

DEFINITION S:rochel eg15 gene.
ACCESSION Y73953

```

ERSION      X73953.1  GI:393391
EXWOPDS     cellbase: cellbase

```

SOURCE Streptomyces rochei.

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetia; Streptomycinae; Streptomycetaceae; Strepto-

REFERENCE 1 (bases 1 to 1470)

TITLE	Direct Submission
JOURNAL	Submitted (07-JUL-1993)
INSTITUTION	Università di Firenze

di Biologia Animale e Genetica, Via Romana, 17/19, 50125 Firenze, ITALY
 2 (bases 1 to 1470)
 Perito, B., Hanhart, E., Irdani, T., Iqbal, M., McCarthy, A. J. and Mastromei, G.
 Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase-encoding gene
 Gene 148 (1), 119-124 (1994)
 JOURNAL MEDLINE 95011642
 FEATURES
 source location/Qualifiers
 1. 1470
 /organism="Streptomyces rochei"
 /strain="A2"
 /db_xref="taxon:1928"
 /clone="E.coli pCSF1"
 90. 103
 /rpt_type=INVERTED
 112. 125
 /rpt_type=INVERTED
 112. 125
 /rpt_type=INVERTED
 152. 157
 /rpt_type=INVERTED
 152. 157
 /rpt_type=INVERTED
 166. 1314
 /gene="eglS"
 166. 1314
 /gene="eglS"
 /FC_number="3.2.1.4"
 /FC_number=1
 /codon_start=1
 /transl_table=11
 /product="cellulase"
 /protein_id="CA52139.1"
 /db_xref="GI:393392"
 /db_xref="SPTREMBL:Q59963"
 /translation="MPRLRHPTLRAVSAALLTALAALATAPAOADTTICEEFGSTVIGRYYVONNRMTSATQCVATGSDGSPVTPNGPKSPYFENFCHYFNSGPTALPARISGISAPSSISYGFVDNAYNASYDIDMPTRGVNTETIMIWFNRGQIOPIGDSQVGTASVAGRTMEVWGSGNGTNDVLSPVAPASMSFVMDPVRATVARGIAGNDMYLTSTIOAGFEPMONGALAVNSSTVNTGSSQRPDPNGDPTPRACTVYATNNVWPGFTNAVITYTNNNGSPVDMKRLAFTLPSQSVVHAMNSVSPSSGAVTATGPASNAITAAIGSSQFPGQAYSGSFAPALFQNLGTACTV"

stem_loop 1321. 1340
 stem_loop 1351. 1403
 stem_loop 1419. 1453
 BASE COUNT 211 a 579 c 475 g 205 t
 ORIGIN

Query Match 12.2%; Score 96.2; DB 1; Length 1470;
 Best local Similarity 47.3%; Pred. No. 4.5e-05;
 Matches 327; Conservative 0; Mismatches 358; Indels 6; Gaps 1;

QY 92 cggagcctgagcccgagccgacgctgctgagcagcgtggaacgagcgagctgtg 151
 DB 254 CCGGACCGCCCGCCGACAGCCGACACACAGATCTCGAAGAGTTCGGCTCGACCGCTCA 313
 QY 152 cgggggggagctacggggtatcaacaagctatggggcgaggaacggccagtgatcg 211
 DB 314 TCCAGGGGCTGTACGTTCAGTAACCAACCGCTGGGACACAGCGCCAGAGCTCA 373
 QY 212 agctgcagcaggaacgaggaacttcagatcagatcagcagggcgatcagcaacaaggacaca 271
 DB 374 CCGGACCGGACTCCGGCTTCGGGTTCACGACGAGCCGACGGCTCGGTCCGACCAACGGCG 433
 QY 272 acgtggcgcctacccgacatctactctgggtgagcacttgaggcgccctgacagagcaat 331
 DB 434 CGCGGAGGTGACCGCTCGGTCTTCAACGGCTGTCTACACGACTGTTCGCGCGGACA 493
 QY 332 cgggagtgccggcgagcggtgagagagctgtccagcagtgagcagagctgagctacgc 391
 DB 494 CCGGCTTCCTCCGCGGATCAGCGGCACTCTCTCCGCGCCAGAGCATCTCTGACGGCT 553
 QY 392 cgaatcaagcagggcgagtggaatgcgcgctacagacatcggtcagtcaccgctcaagaat 451
 DB 554 TCGTGACACACCGCGGTACACAGCGCTGTACGACGATCTGGCT-----GGACCCGACGC 607

QY 452 ccggcaacggtctacagcgcgccgagcgtgatactgtgctgaacttgaaacgagcgagc 511
 DB 608 CCGGACCGGACGGGTGACACCGGACGAGATCATGATCTGTCTCAACCGGCTGGGACAGA 667
 QY 512 tgatgccggcgagcgagcgagcggtgagcagcggtgaaacttgagcggggcgacatggagctc 571
 DB 668 TCCAGCGGATCGGCTCGACAGTCCGACGGCCGCCGCGGGCGGACCTGGGAGGTGT 727
 QY 572 ggtatgccagctggaacttgaaatcaatcgcctaccgagcgagcaagaccccaagctcg 631
 DB 728 GGTCCGGCGGGAACGCGACACGACGACTGCTGTCTGTCGCTCCGTCGCGCATGAGCA 787
 QY 632 tgaagcagctggaacttgaaagccttcaatcgaacgagcgagtcgcgcgagctacatccgc 691
 DB 788 GCTGAGACTTCGACAGTGATGATGATCGTCCGGGCGACCGTCGCGCGGCTGGGAGCA 847
 QY 692 cggagtggtatctgcatgctggtgagagcggttcgaactctggaggggggcgagctg 751
 DB 848 ACGACTGCTACCTGACGACGATCCAGGCGCGTTTCGAGCCCTGGCAGAACGGTGGGAGTC 907
 QY 752 tgcgaagcgccgatttccgtaacggtgca 782
 DB 908 TCGGCGTCACTCTCTCTGTCGACCGTGAA 938

RESULT 10

AF130408 1311 bp mRNA linear BCF 14-APR-1999
 LOCUS Streptomyces viridosporus strain T7A cellulase (celS1) mRNA,
 DEFINITION complete cds.

ACCESSION AF130408
 VERSION AF130408.1 GI:4583444

SOURCE Streptomyces viridosporus.
 ORGANISM Streptomyces viridosporus.

REFERENCE 1 (bases 1 to 1311)
 Ramachandran, S. and Crawford, D. L.
 Characterization and sequence analysis of two genes involved in cellulase degradation in Streptomyces viridosporus T7A, and its expression in Escherichia coli

AUTHORS Ramachandran, S. and Crawford, D. L.
 TITLE Characterization and sequence analysis of two genes involved in cellulase degradation in Streptomyces viridosporus T7A, and its expression in Escherichia coli

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1311)
 Ramachandran, S. and Crawford, D. L.
 Direct Submision
 Submitted (24-FEB-1999) Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, LSN 125, Gibb Hall, Moscow, ID 83844-3052, USA

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1311)
 Ramachandran, S. and Crawford, D. L.
 Direct Submision
 Submitted (24-FEB-1999) Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, LSN 125, Gibb Hall, Moscow, ID 83844-3052, USA

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1311)
 Ramachandran, S. and Crawford, D. L.
 Direct Submision
 Submitted (24-FEB-1999) Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, LSN 125, Gibb Hall, Moscow, ID 83844-3052, USA

FEATURES

source

1. 1311
 location/Qualifiers

/organism="Streptomyces viridosporus"
 /strain="T7A"

/db_xref="taxon:67581"
 1. 1311

/gene="celS1"
 82. 97

/function="putative regulatory region"
 /rpt_type=inverted
 144. 149

/gene="celS1"
 157. 1311

/note="putative"
 157. 1311

/gene="celS1"
 /note="celS1; family H member"

/codon_start=1
 /transl_table=11

/product="cellulase"
 /protein_id="AD25090.1"

/db_xref="GI:4583445"
 /translation="MRRLAPDAPCAVMSPPSPSPSSRRMRHRAOADTTICEP

YGTITIGRYYVONNRMTSATQCVATGSDGSPVTPNGPKSPYFENFCH

YFNSGPTALPARISGISAPSSISYGFVDNAYNASYDIDMPTRGVNTETIMI

WFNRGQIOPIGDSQVGTASVAGRTMEVWGSGNGTNDVLSPVAPASMSFVMDPVR

ATVARGIAGNDMYLTSTIOAGFEPMONGALAVNSSTVNTGSSQRPDPNGDPTP

PRACTVYATNNVWPGFTNAVITYTNNNGSPVDMKRLAFTLPSQSVVHAMNSVSPSS

GAVTATGPASNAITAAIGSSQFPGQAYSGSFAPALFQNLGTACTV"

[illegible]

correct initiation codon. Where possible we choose an initiation codon (atg, gty, ttg or (atc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid G114 lies on the AseI-g genomic restriction fragment.

```

LINES
source
1. .41782
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid F37"
1. .1299
/gene="xylB"
1. .101
/gene="xylB"
/note="nominal overlap with Streptomyces coelicolor cosmid 2Stc11"
CDS
<1. .1299
/gene="xylB"
/note="SCG11A.01, xylB, xylose kinase, len: 432 aa; highly similar to SW:XYLR_STRRU (EMBL:M73789) Streptomyces rubiginosus xylose kinase (EC 2.7.1.17) (xylosekinase) xylB, 481 aa; fasta scores: opt: 2553 z-score: 2795.9 E()): 0.89, 2% identity in 427 aa overlap. Contains match to Pfam entry PF00370 FGCV, FGCV family of carbohydrate kinases and two Prosite matches to entries PS00933 FGCV family of carbohydrate kinases signature and PS00445 FGCV family of carbohydrate kinases signature 2"
/codon_start=1
/transl_table=11
/product="xylose kinase"
/protein_id="CAB61582.1"
/db_xref="GI:646842"
/translaction="DPROMMDALGEGALSCQGEAREAAAVSVGGQOGLVTLDRGGEFVAPLALMMDVSPAPORRLIDELGGAKMAEETGSPASFTVTKMALTDEEPAEA AVKAVRLPHDVTYTERLTGEGTDRGDVSGTGMWASGTAVDEILARALDALLPRVARGVACTVRKGHGLPFSKGTLYVAGCTGDNRNAAAGLGRPGVPVMSLGTSGRATAVSQRADPTGTVAGFADARGDWPLACTLNLCTLAVDRAVSLGLDREAVEGTDTLLPFLDERPNTLPHSSGLHLGRLHDTAGQLQAAVDGAVHSLGLDLVDADPDSPLLTLGGCARGAOQTVARLRSGRPVQIPPEARLVALGAQAAGLTLGEDAAVARRWRTAGPVDAVERDEATLNRITGVLSDAAPLTERDAASR"
76. .1122
/gene="xylB"
/note="Pfam match to entry PF00370 FGCV, FGCV family of carbohydrate kinases, score 514.30, E-value 8.7e-151"
250. .288
/gene="xylB"
/note="PS00933 FGCV family of carbohydrate kinases signature 1"
76. .936
/gene="xylB"
/note="PS00445 FGCV family of carbohydrate kinases signature 2"
1519. .1523
1525. .2733
/gene="SCG11A.02"
1525. .2733
CDS
/gene="SCG11A.02"
/note="SCG11A.02, possible xylose repressor, len: 402 aa; similar to SW:XYLR_BACSU (EMBL:M27248) Bacillus subtilis xylose repressor xylR, 384 aa; fasta scores: opt: 612 z-score: 642.4 E(): 2.1e-28; 29.4% identity in 381 aa overlap. Contains match to Pfam entry PF00480 ROK, ROK family and a possible helix-turn-helix motif at residues 43. .64 (+3.15 SD)"
/codon_start=1
/transl_table=11
/product="putative xylose repressor"

```


Query Match

```

/protein_id="CAB61583.1"
/db_xref="GI:6468423"
/translation="MSAPPHENQAPRPAGALPDTOQGMRRRNLSRMHTVSAEGPLSR
AASVRLITRAAVSTIDELIRLSGLEELGEPERGVSRPGSALVSGGPGAGIE
GVGSHLACAVDALRQVRVRAVPMIDMHADLGLALPADLPETVNEAFGALATMLGD
GLAAVPELVASDGTIGGAANVEFGRLIRGRGAGELGHPVPHVDGRCACGGGCELD
ZPRPHELVASDGTIGGAANVEFGRLIRGRGAGELGHPVPHVDGRCACGGGCELD
YAGVAGVLAARAAGEGDEBRVGLAGRAAGCDVDRALLRAGATLIALTGAVNLIDP
EYGVGGLAGLACIAPLRLPSRLBELRAATGRACPVASVSELGPGPLGAHNVRAVL
DDPGAVASERA"
1822..2373
/gene="SCG11A.02"
/note="Pfam match to entry PF00480 ROK, ROK family, score
131.80, E-value 4e-38"
3041..3646
/gene="SCG11A.03"
3041..3646
/gene="SCG11A.03"
/note="SCG11A.03", probable amidase, len: 201 aa; similar
to Sw:AMPD_ECOLI1 (EMBL:X15237) Escherichia coli
anhyd-N-acetylmuramyl-tripectate amidase AmpD, 183aa;
fasta scores: opt: 211 z-score: 267.1 E(): 1.7e-07; 39.5a
identity in 114 aa overlap and to Streptomyces coelicolor
SC2A11.21c; fasta scores: opt: 674 z-score: 751.6 E(): 0;
50.0a identity in 202 aa overlap. Contains match to Pfam
entry PF01510 Amidase-2, N-acetylmuramyl-L-alanine
amidase"
/codon_start=1
/translation="11
/product="putative amidase"
/protein_id="CAB61584.1"
/db_xref="GI:6468424"
/translation="MERETAPVPRRRLKKAALATVPTYLSTGRAAAQVRAVDY
SAEMIPACSTSNRSRSPRAVPEVDVYVAVTQETADLISIFRDPKQSAHVRS
DGHNAACREDDIENAHAGWMDVNTSISGHEGMDRPEFTNAMYEQSARLTAIT
AYGPRKDRTHIIAHNEHVEPGSDHDPGPFMDWRYRLNVNA"
3191..3637
/gene="SCG11A.03"
/note="Pfam match to entry PF01510 Amidase-2,
N-acetylmuramyl-L-alanine amidase, score 89.80, E-value
5.7e-23"
3821..5080
3821..5080
/gene="SCG11A.04"
3821..5080
/gene="SCG11A.04"
/note="SCG11A.04", possible transcriptional regulator, len:
aa; similar to TR_031551 (EMBL:299108) Bacillus subtilis
transcriptional regulator ACR, 605 aa; fasta scores: opt:
393 z-score: 446.5 E(): 1.7e-17; 34.7% identity in 199 aa
overlap"
/codon_start=1
/translation="11
/product="putative transcriptional regulator"
/protein_id="CAB61585.1"
/db_xref="GI:6468425"
/translation="MTDWMALIEPQAPVERAPARLBRHNETTEAGTVRPPVAAVAD
SMRSVRAGVEDDGAISVEMLDGAYAEKPLSVMLRRELLSTFRAPDGEHLAAV
CDAGKRLIWEBSHPRTRRAGRMNVPGAKWSESAGVNAFGVAVVGRVQVFAEEN
TIRRAVQPTCAAAVPAHDPTGVGADVDTGGDGLAHPSLISVQAVARAESQIALLL
TPERSASAAETALIGDREGLISADGRVRLRMRSEIIVLLAHNEGPTGDBLLAL
YEDVEVPTVLAELARLIGRGLASRPYRLRHWSESDTSVNERLRACAVTAAV
AAVAGPLIPQAPVGRRLRRRLDGLRALIACGPDLDADNAHAWGSDGDDVWRA
LAAVRPTATSRRLAALESELAPRW"
5205..5209
5218..6741
/gene="hchA"
5218..6741
/gene="hchA"
/note="SCG11A.05, hchA, aldehyde dehydrogenase, len:534
aa; highly similar to Sw:THCA_RHOSN (EMBL:017129)
11.1%; Score 87.6; DB 1; Length 41782;

```

Matches	298:	Conservative	0:	Mismatches	324:	Indels	6:	Gaps	1:
Best Local Similarity 47.5%; Pred. No. 0.00033;									
Oy	156	gggagcctacccggtgatcaaacgta	tg	gggagcggagagacggccagtgcat	ga	gt	215		
Db	27828	ggggcagatgacgctgcgacgaac	aa	ccgctggggctccaccgacccagtc	gctac	cgcc	27887		
Oy	216	cggactgtgaaagggcgaacttca	gc	atacaacagggcgtatcaacggca	acgg	caacaa	275		
Db	27888	caccgacacacggccttcgggtr	ta	ccagccagccgcttcggcaccgac	caac	aaacggggcc	27947		
Oy	276	ggccgctcatcccgccatctact	tg	gggtgcacatggggcgcttgcag	ga	gaatctgg	335		
Db	27948	gaagctgatacccgctggcttca	ag	ggctgcacatgacgaactgtttc	ac	ggcgccg	28007		
Oy	336	attgcgcggcgcggtgacggag	ct	gtccgaagttggcagcagggcttg	aa	ggcccaagc	395		
Db	28008	cctcccccgtccggctgcgaac	ac	gcttccggccgcttcacacatct	gta	ggccttgc	28067		
Oy	396	caacgacgggcgctgtgaaatg	cc	gctacagacatctgtgtacgtcc	gctac	gaattccg	455		
Db	28068	cgacggcgccgcttccacagcc	ag	ccgacggcttcggcaccgacac	caac	aaacggggcc	28121		
Oy	456	caacgctacacagcggcgccg	cc	gagctgtatctgctgtgaacttg	aa	cgccgctg	515		
Db	28122	caccacgggggtgaacacga	ac	cgagatcatgtgtggttcaac	ag	gggtggtccg	28181		
Oy	516	gccggcgccagccgctgycgc	ac	cgcttgaaatggccggatcgta	gta	575			
Db	28182	gcccatccgcttaccgggtgg	ac	ggcgccctccgcccggaccta	ctgg	ag	28241		
Oy	576	tgcgacgtggagctggaatt	ta	atcgtcctaccggcgacga	agcc	accacgtctg	635		
Db	28242	cggccggcaacggccttgcga	ac	gagctcgtctgctgacccgtg	cgga	28301			
Oy	636	cgagcttgacacgaagccttc	at	cgacaacggcgcgcctacatc	ccg	cgga	695		
Db	28302	gagcttgcacgctgacgactt	gc	tcggcgccgacacgctccgc	cgac	ta	28361		
Oy	696	gtgtatctgtcatgcggtgtg	aa	cgggcttcgaactctggaggg	ggcg	cgctctg	755		
Db	28362	ctggtaacctgacgacgctt	ac	ggcggttcgacccgtgcag	ac	cgccgactg	28421		
Oy	756	aagcgcgatcttccgtaag	gt	cg	783				
Db	28422	cgtgacatcttctctccac	cg	ctgac	28449				
RESULT 13									
LOCUS	AE005123	11202 bp	DNA	linear	BCT 12-FEB-2001				
DEFINITION	Halobacterium sp. NRC-1	section 154	of 170	of the complete genome.					
ACCESSION	AE005123	AE004437							
VERSION	AE005123.1	GI:10581826							
KEYWORDS									
SOURCE									
ORGANISM	Halobacterium sp. NRC-1.								
	Halobacterium sp. NRC-1								
	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;								
	Halobacteriaceae; Halobacterium.								
REFERENCE	1 (bases 1 to 11202)								
AUTHORS	Ng M.W., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M., Shkrel H.D., Lasky S.R., Baliga N., Thorsson V., Sbrogna J., Swirell S., Weir D., Hall J., Dahl T.A., Welti R., Gao Y.A., Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W., Maddocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H., Isenbarger T.A., Peck R.F., Pohlschrod M., Spidlich J.L., Jung R.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omig A.D., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L. and DasSarma S.								
TITLE	From the cover: genome sequence of halobacterium species NRC-1								
JOURNAL	Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)								
PUBMED	11016950								
REFERENCE	2 (bases 1 to 11202)								

1670 a 4414 c 3626 g 1492 t
 CDS
 gene
 /translation="MLSALVADHIITHTKNGFAFTTLMGLYALLIOGSPGRLSDLSRK
 TLIVTSLGASGGFVNVATARSVALITVGAIVTGGTGLPTARALVADLVSTKOR
 ARFGHTPAGSDLGCHSANGIAVVALALLSMRRAPVPIYCLTAAVAIALHWRSREYVE
 RRSALVADTARLLSTGQFRWILVATYALFTWQAIAPLFTFTDSKEPAVPGVVA
 FGALFVAGAIAPKPTAGTSDTISRRAALVAGALVGAALSAVVVPSVPVAGVAFV
 AAGLAFPPVWQSYLMDAPFPESAGGDLGAMRTVYIGLGLGPTYGVASTAANYEIA
 FMCGLVGLIVAAVVIARVT"
 complement(9890..11137)
 /gene="thrc1"
 /note="VNC2430G"
 complement(9890..11137)
 /gene="thrc1"
 /note="thrc1"
 /codon_start=1
 /product="threonine synthase"
 /transl_table=1
 /protein_id="AAG20514.1"
 /db_xref="GI:10581834"
 /translation="MTDLSLAGETNPACATDGGVLMACEVCGGTFLAPADVYVRCPECSG
 LLEARYADYPGFEAFSGRGVWYRVEDALVPDAGVSQVGDTPLELAAVPAAGAAGVADVR
 VKHGEMPTGSGDRGMTGVYRVARERLGVORLACASGNTSALPAAGAAADVPVVLV
 LPAKVAAGKIAQALAGARICEVNGDFEDLDVVALARGAEYILNLSINPFLBEO
 KTIKFLLEOSRRTATGEMPRIVLPVNCACNTALYAFREIYVAAAGMTDEMDPALNG
 VQAAGAPMVEAVACGRTERRTVDVETRTATITIGRPVNAKPLPVEYETGGTAAV
 SDDITDQRMLAGDGIIVGEPSAASVAGLLALEREKEISADERVCLTTGHLIKDPA
 AAAAAGSDTTPVPDAGCVLDAL"

[illegible][illegible]


```
RLAROPARNDKVVTSWNGLAITLALAEASVALDDPALAHARCATRLIDLHVNDRL
RASLSGGVVDSSAILEDHAMLATGLALYQLTSEGAWLTAATGLDITVAHGDQR
PGRWEDTADDADRLRPSDPLDGAIPSGASSIAEMLTAGHVGDGARERWOLAD
TLRAHVLILARAPRSRGMWLAVEAVRPLQIIVACDIPRSSLLADARRLRAPGCAIV
VGAAGSALLVRDRBVAGADAAVACRGVCDLPVTSAEMLATGALGVP"
4429. .4548
/gene="MT1116"
4429. .4548
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45372.1"
/db_xref="GI:13880697"
/translation="MAPVHREIRDAVSACRAIFLTALALGVPGQNTNHIAH"
complement(4518. .5246)
/gene="MT1117"
complement(4518. .5246)
/note="similar to PID:887849 GB:U00006 PID:1789266;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Hly-III family protein"
/protein_id="AAK45373.1"
/db_xref="GI:13880698"
/translation="MSGQDPTATTAETAPPAHAHHLVEGVARVLTKEPRFGMIHYVS
ACTAVLAGASLVAWSVAGSARAGLTTLATTAITTFMVSATYHHRNKSADARMM
KRADSMITFVFAGSYTPRALALPAHGRVLSLYWGALISGILYAVRMPDPTTFGY
VPLYLIGMVAWYMTRTIILHNGVTLVLFVGALISIGILYAVRMPDPTTFGY
HEFHACTVAAVACHIAMVEVF"
5357. .6145
/gene="MT1118"
5357. .6145
/note="similar to SP:P20182 PID:1196916; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="undecaprenyl diphosphate synthase"
/protein_id="AAK45374.1"
/db_xref="GI:13880699"
/translation="MTIPPRLEKEPLYRIEYRLRQGLAASDLEPRHIAVLCGDRNR
WARSAGYDVSYGIRMGAAKIAEMLMCHGAGIELATVLLSTENLQRPDELALIE
ITDVAEEICANMWSVRYVDLGLIGEPARLRGAVESTPEVASFPVNAVAGYG
REIIVDAVRLSKELANGATAEELVDATVGEISNELYTSQPDPLIVIRTSGBQRL
SGELMOSAVSEMWTFEAMWPAFRHVDPLRALRDSARHRSYGR"
6283. .8591
/gene="MT1118.1"
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
8768. .9088
/gene="MT1118.2"
8768. .9088
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45375.1"
/db_xref="GI:13880700"
/translation="MPCVGYGDRREFVAVAVEAICENLNTSGQDPDLVIRTSGROR
LSGHRGPTGVSRRRLRLALRDYSTPHASIPYPPYRSDGIHASRLAVESFDALAG
RYEL"
9223. .9675
/gene="MT1119"
9223. .9675
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
```

```
/product="PE family protein"
/protein_id="AAK45376.1"
/db_xref="GI:13880701"
/translation="MGKGLRMSYMLATPALTAATDIDIGSAVSANAAVAATTG
VLAAGDEVILAIARLFLNMAEYHLSNOVAFOPLFRLTGGCGFRRRRGRQCV
TAEHRAAGAGRRQRRRRSGDQWMLRQRRHFGCGGCPFRHSHRR"
9692. .9859
/gene="MT1120"
9692. .9859
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="PE family protein"
/protein_id="AAK45377.1"
/db_xref="GI:13880702"
/translation="MTTASATASSTGVDEGIATYAVASQMDGVVAVNTINQERDF
```

```
Query Match 8.2%; Score 64.4; DB 1; Length 14483;
Best Local Similarity 46.8%; Pred. No. 0.99;
Matches 307; Conservative 0; Mismatches 341; Indels 8; Gaps 3;

QY 132 ctggagacgcgcgcatgtgscggyggygcctacccggtgatcaacaagatagggc-g 190
DB 10132 CTAGTTGGGAGCAGCTACGCTCCAGGAGCGTACGTGCTCAGAACACGCGTGAACAA 10191

QY 191 cggagagccgcgcagtgcatgtaagtgagactggaacaggcaactcacatcacacggg 250
DB 10192 CCCCCCGCGGCGACCGCTCAACGTCAAGCAACCCGGTTCCACCATCACCGAATGAAG 10251

QY 251 ccgaatacgaacaaaggcaacaacgctggccgctatccgacatcactcctcggctgaact 310
DB 10252 GTGCTGCCCCCAACCAACGGCGCCCGCTGAGTTAACCCCTGATGCTCGAGGGCTCCACT 10311

QY 311 gggg-cgacctgacaggaacttcggatgacgscgscgscggtgacgagagctgtccagcgtc 369
DB 10312 GGGGCGACCTCGTGGTGGGCGACCAACTGCTCTACTGAGGTGGGCCAGATTGTGGGGC 10371

QY 370 cgcacgagctgagcgtcgaacatccgaatacgaacgagccgctggaatgcccctacgaacatc 429
DB 10372 CGACACGAGCATTCAGTACACATCAACCGAGACCGGGGTATGGAGCGCTCTCGACATTC 10431

QY 430 tggctcaagtcgctcaagaaatccggaacgctacagcggcgcgcgagctgatgatc 489
DB 10432 TGCCTGGATTCCACACCACCAAGACGACCGGGGTCAACC-----AGCAGGAGATCATGATC 10485

QY 490 tggctgaactggaacggcgcgctgctgacgagccgagccgctggccaccgctggaactc 549
DB 10486 TGGTTCAACCAACCAAGGCTCTCAATTCAACCGGCTGGCTCCCGGTGGGCAACCAACCATTC 10545

QY 550 gccgggcccacctgggaagctcgtatgctcgaactgggaactgggaactacatcgcctacgg 609
DB 10546 GAGGCGAAGAACTTTCGTGTGTGGATGGCAGCAAGGCGATGAACAACCGGATGGCTTAT 10605

QY 610 cgcagcagcccaacacgctggtgtagcgagctggaactgaagcctcatcgacgaacgg 669
DB 10606 GTGCGCAGCCAGCGCATTCGAGCTGTGAGCTTCAGCTGATGATTTTCGCGACCAACC 10665

QY 670 gtgcgcggcggtacatccgscgagtggtatcgtatcgatcggttggaagcggcttcgaa 729
DB 10666 GCCACCATGAGGCGCATTCACCGACTGTGTACTCAACGACATCCGGGCGGCTTGGAG 10725

QY 730 ctctggagggcgagggcgctcgtcgaagcggcagatcttctcgtlaacggtgcagta 785
DB 10726 CCCTGGAGCAGCGTGTGGGTCTGGGGGTGATTCGTCTCGCGCGAAGTCAACTA 10781
```

Search completed: August 19, 2002, 12:44:44
Job time: 10428 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:37:08 ; Search time 544 Seconds
(without alignments)
2480.689 Million cell updates/sec

Title: us-10-003-759-3

Perfect score: 786
Sequence: 1 atgaacgtcatcgtcgtcgtt.....tttcgcgaacgttcagtag 786

Scoring table: IDENTITY_NUC
Gapop 10-0, Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	131.8	16.8	1116	20 AAV72359
2	131.8	16.8	1116	21 AA257029
3	131.6	16.7	1715	21 AA257031
4	57.6	7.3	1983	22 AA454513
5	53.6	6.8	77536	21 AA414651
6	49.8	6.3	756	17 AA16766
7	49.4	6.3	1404	21 AA38763
8	49	6.2	756	14 AA043032
9	48.8	6.2	985	19 AA64548

10	48.8	6.2	985	19 AA44439
11	48.8	6.2	985	20 AA219349
12	48.8	6.2	985	20 AA219137
13	48.4	6.2	12152	22 AA508699
14	48	6.1	4403765	22 AA199683
15	47.8	6.1	24379	18 AA793095
16	47.8	6.1	24379	19 AA25925
17	47.6	6.1	30001	18 AA761016
18	47.6	6.1	30001	20 AA505110
19	47.2	6.0	1908	16 AA099364
20	47.2	6.0	1908	20 AA584065
21	47.2	6.0	1908	21 AA661403
22	46.8	6.0	15872	21 AA287283
23	46.8	6.0	20394	22 AA724892
24	46.6	5.9	2712	19 AA794214
25	46	5.9	44377	18 AA78508
26	46	5.9	44377	18 AA780414
27	45.6	5.8	744	18 AA774073
28	45.2	5.8	1291	20 AA87940
29	45.2	5.8	2852	19 AA64558
30	45.2	5.8	2852	19 AA219359
31	45.2	5.8	2852	20 AA219147
32	45.2	5.8	2852	20 AA219147
33	45.2	5.8	77536	21 AA414651
34	44.8	5.7	2268	21 AA410501
35	44.8	5.7	2618	21 AA38760
36	44.8	5.7	3252	21 AA410505
37	44.6	5.7	2376	20 AA416153
38	44.6	5.7	5324	22 AAC90079
39	44.6	5.7	11220	21 AA287298
40	44.6	5.7	12588	15 AA063293
41	44.6	5.7	36778	21 AA287318
42	44.6	5.7	37948	21 AA287285
43	44.6	5.7	38506	21 AA475633
44	44.6	5.7	38506	21 AA256001
45	44.4	5.6	867	22 AA486828

ALIGNMENTS

AAV72359	1	AAV72359 standard; DNA; 1116 BP.
XX	XX	AAV72359;
AC	XX	28-JUL-1999 (first entry)
XX	XX	Actinomyces sp. 36kD cellulase DNA.
DE	XX	Cellulase; detergent; animal feed; nutritional value; textile;
KW	XX	stone washing; texture modification; appearance; cellulosic fabric;
KW	XX	pulp; draining; paper; baking additive; starch treatment; grain;
KW	XX	high-fructose corn syrup production; ethanol production; fibre reduction;
XX	XX	milling; ss.
OS	XX	Actinomyces sp.
XX	XX	WO9925846-A2.
PN	XX	27-MAY-1999.
PD	XX	18-NOV-1998; 98WO-US24649.
PF	XX	24-JUN-1998; 98US-0104308.
PR	XX	19-NOV-1997; 97US-0974041.
PR	XX	19-NOV-1997; 97US-0974042.
PA	XX	(GENV) GENENCOR INT INC.
XX	XX	Jones BB, Van Der Kleij WAH, Van Solingen P, Weyler W;
PI	XX	

Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
M. carbonacea DNA
Mycobacterium tube
Streptomyces feno
Streptomyces roseo
Total DNA sequence
S. aureofaciens DN
S. lividans protea
Tripeptidyl aminop
CDNA sequence enco
S. venezuelae vep
Pimaricin biosynth
Polyangium brachys
Platanoides syntha
Platanoides syntha
Myceliophthora the
Mycobacterium tube
M. tuberculosis im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Nucleotide sequenc
Non-reducing sacch
Rice 1-deoxy-D-xy
Non-reducing sacch
Mouse Sox1 CDNA.
L05390 CDNA clone.
S. venezuelae macr
Sequence encoding
S. venezuelae pik
S. venezuelae pik
Nucleotide sequenc
Recombinant cosmid
H. vitrescens inorg


```

Db 257 agtctatccctcgtctgaagcagcgctgcacactgagcagcccgacgagc 316
Qy 338 tgcgcgcgcgtgcaagactgtcgcagcgtgcagactggaagctcagcgagc 397
Db 317 tgcctatgctgcatcagctcagcagcgccgagcagctgtctctacgctacacg 376
Qy 398 cgaagcgctggaatgcgcgtacacatctgttcaagtcctccgtacgaatccgca 457
Db 377 gcaacgctgtacaaacgcgcgtacgaatcttg-----ctgagccgacaccccgca 430
Qy 458 acggtctacagcgcgcgagcagctgtatgtctgtcgtcgaacttggaacgcggtatgc 517
Db 431 ccaacgggtgtaacccgagcatctgtctgttcaacccgggtcgcggttcagc 490
Qy 518 cgggacgacgcgcgtgcaacccgtgaactgagccggcgagcagctgggaagctgtatg 577
Db 491 ccatcggttcgcggttcgcaacgcgcacgtcggcgcgccagctggaagtggtgacgc 550
Qy 578 cgaactgggaactgaatatacgtcctacgcgcgacgacccacacgctcggtgagcg 637
Db 551 gcagcaacggttcgaacagctgtatctctctcgtgcgcctccgcatcagcagctgga 610
Qy 638 agctgaacccgaagcgcttatacgaacgagctgcgcgcgcgtacatccgcgcgaat 697
Db 611 gcttcgacgtcaagagactcgtcgaacgagcgcagcgcgtgagccaccccgagct 670
Qy 698 ggtatctgcatgctggtgacgagcgttcgaacttggaagggcgcggtcgtcgaa 757
Db 671 ggtacccacacagcatcgaagcggtctcgaacgctggaagggcgagcagcggtcgtgcg 730
Qy 758 ggcgcgacttttcgtaacggtgca 782
Db 731 tgaactcgtctctccgcggtgaa 755

RESULT 3
AA257031 ID AA257031 standard; DNA: 1715 BP.
XX AC AA257031:
DT 19-MAY-2000 (first entry)
XX DE DNA sequence of an expression cassette containing cellulase 11AG8.
XX KM Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
XX pulp; paper; cellulase 11AG8; ds.
OS Streptomyces lividans.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT promoter 1..406
XX FT /tag- a'
XX FT /note- "GI promoter sequence"
XX FT sig_peptide 407..548
XX FT /tag- b
XX FT /note- "Cela signal sequence"
XX FT mat_peptide 549..1564
XX FT /tag- c
XX FT /note- "mature 11AG8 sequence"
XX FT misc_feature 1576..1715
XX FT /tag- d
XX FT /note- "11AG3 downstream sequence containing terminator"
XX PN MO200009707-A1.
XX PD 24-FEB-2000.
XX PF 28-MAY-1999; 99MO-US11971.
XX PR 24-JUN-1998; 98US-0104308.

```

```

PR 18-NOV-1998; 98MO-US24649.
PR 28-MAY-1999; 99US-0321981.
XX PA (GENV ) GENENCOR INT INC.
XX PI Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
XX DR WPI: 2000-224344/19.
XX DR P-PSDB; AAY67497.
XX PT A novel Actinomycetes cellulase and related DNA, useful for detergent
XX compositions, treating textiles and paper or pulp
XX PS Example 6; Fig 15; 72pp; English.
XX CC The invention provides a cellulase from Actinomycetes. The cellulase can
XX be used in a detergent composition, as an additive for animal feed and
XX for the treatment of textiles or pulp and paper. The DNA encoding the
XX cellulase can be used to identify homologous cellulases and for
XX recombinant production of cellulases. The present sequence represents
XX the DNA sequence of a complete expression cassette consisting of the GI
XX promoter, cela signal sequence, cellulase 11AG8 and GI terminator
XX SQ Sequence 1715 BP; 306 A; 612 C; 558 G; 239 T; 0 other;

Query Match 16.7%; Score 131.6; DB 21; Length 1715;
Best Local Similarity 50.7%; Pred. No. 1,7e-18;
Matches 345; Conservative 0; Mismatches 329; Indels 6; Gaps 1;

Qy 103 cccgaagcgaacgctcgagctgtcgcgaacgctgggaacgcgcgagctgtgcccggggcgcc 162
Db 533 ccgcaacagggcacaacagcagatctgacgcgtacgacccacacgacatccagacgcg 592
Qy 163 taccgggtgatcaacaacgatatggggcgagagacgcgcgaagtgatgggtgcgactg 222
Db 593 taccgtgtgacagaacaacgcgttggggacacgacgcacccagtgatcaatgtgacccgc 652
Qy 223 gaaacgggcaacttcaacatcagacgggacgatacagacaacggaacagctggccgc 282
Db 653 aacggttcgagatcaaccacgagcggttcggtgcgcgacaaagcgcccggaagctc 712
Qy 283 taccggcctactactcgtgtgcacactggggcgctgcagagcaatctggatggcg 342
Db 713 tatccctcgttctacgagcggttcacactacgcaactgcgcgccgcgacgacgtgtccc 772
Qy 343 cggcgcggtgaggaagcgttcgcagcggtgcgacgagctggcgctcgaacgcgacagc 402
Db 773 atcgagatcagctcgtatcgcagcgcccgacagtggtctctctacgctacacccggcaac 832
Qy 403 ggcgcgtggaatgcgcgtcagcatctgtgtcagtcgcgtacagaaatccggcacaacgc 462
Db 833 ggcgtctcaacagcgcggtacgacatctg-----ctgagcccgaaaccccgacacaac 886
Qy 463 tacagcgcgcgcgcgagcgtgatgatctgtgctgaacttggaacggcggtgtatgcgggc 522
Db 887 ggggtggaacggacagatcatgatctgttcaacgggttcggccgggtccagccatc 946
Qy 523 ggcagccgctgacacagctggaacttgcggggggccacctgggaagctgtatgcgac 582
Db 947 ggttcgcggttcgcagcgccacagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1006
Qy 583 tgggaactgaatatacgtcctacccggtcgcagcagccacccacgltcgtgagcgagctg 642
Db 1007 aacggttcgaacgagtgatctctctctcgtgcgcctccgcatcagcagctgagcttc 1066
Qy 643 gacctgaagccttcatcgcagcgcggttcgcccgcggtcacaatccggccgggtgtgat 702
Db 1067 gacgtcaagactctgcagccagcgcgctcagcagcgcccttgccaccccgagctgtac 1126
Qy 703 ctgcatcggttggaacgagcgttcgaactctgggaagggcgggcgcggtctcggaagcgcc 762
Db 1127 ctcaacgacatcgaagcggtctcgaacgctggaagggcggaacggtctcgtgcgctgaaac 1186

```

QY 763 gattttccgttaacggtgca 782
 || ||| ||||| |
 Db 1187 tegtctctctccgcggtgaa 1206

RESULT 4

AAA54513
 ID AAA54513 standard; DNA; 1983 BP.

AC AAA54513;

DT 11-APR-2001 (first entry)

DE Fructan exohydrolase (FEH) coding sequence.

KM Fructan exohydrolase; FEH; transgenic plant; recombination;

KW transgene; gene expression; detergent; detergent additive;

KX oral care composition; ds.

OS Zea mays.

PN WO200068402-A1.

PD 16-NOV-2000.

PF 08-MAY-2000; 2000WO-EP04226.

PR 06-MAY-1999; 99BE-0000329.

PI (LEUV-) LEUVEN RES & DEV.

PI Van Den Ende W, Van Laere A, De Roover J, Michiels A;

DR MPI; 2001-007401/01.

PT Novel DNA molecules encoding enzymes having fructan exohydrolase

PS compositions, and in detergents

PS Claim 9; Fig 4; 45pp; English.

CC Transgenic plants such as Cichorium litybus, Cynara scolymus,

CC Helianthus tuberosus, Scorzonera hispanica, Oryza sativa, Zea mays,

CC Trifolium aestivum, Trifolium durum, Hordeum vulgare, Secale cereale,

CC Avena sativa, Sorghum vulgare, Phleum pratense, Lolium temulentum,

CC Dactylis glomerata, Pennisetum americanum, Allium cepa, Agave

CC americanum, Agave azul tequilana, Sorghum bicolor and Panicum milaceum,

CC transformed with a vector encoding a fructan exohydrolase (FEH)

CC enzyme are useful for the recombinant production of FEH or other

CC polypeptides having FEH activity. The FEH polypeptides produced

CC are useful in detergents or as a detergent additive and in oral

CC care compositions.

CC Sequence 1983 BP; 357 A; 614 C; 683 G; 329 T; 0 other;

QY 151 gccggggggcgctaccggtgatcaacaacgtatggggcgcgagagaccgccagtcaat 210

Db 796 gccgggggtgcgacgctgctgaagctcagcgatgatgacacgctcagaccatactg 855

QY 211 gaggtcgtgactgaaacggcaactcagatcaacgagcgacgacgacgacgacgac 270

Db 856 gtggggcggtacgacgacgagcggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 915

QY 271 aacgttgcgcgcctatcctacttctcgttgcgaacttgcgcgcgcgcgcgcgcgcgcgc 330

Db 916 gacgttcgcggtcgc 972

QY 331 tcggatttcgcgcgcgcgttcgagagactgtccagac-gtgcgcacgagctggacgtctac 389

Db 973 ttcgaagcgcgcgagcgagcggtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1032

QY 390 gccgatacgaagcgagcggtgtgaaatgcgcgtacgacatctgttcaatccgtcagaa 449

Db 1033 gccgacgacgtcgcgaaggtctgttcgcgcacgttccgcgcgaagctgtgtcgtcagag 1092

QY 450 ttcggcaagcgcttaagcgcgagc--gccgagctatgatctgtcgtgaacttgaacgcgc 507

Db 1093 gacgggaagcagctcgcgcgcgtgtgcgtgtgagagatagaagaacgtcgcgaggaagcgc 1152

QY 508 gccgtatgc 565

Db 1153 gtgcgtctcgc 1210

RESULT 5

AAA14651/c
 ID AAA14651 standard; DNA; 77536 BP.

AC AAA14651;

DT 08-AUG-2000 (first entry)

DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.

KM FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;

KM Streptomyces hygroscopicus var. ascomyceticus; immunophilin;

KM FK-506 binding protein; polyketide compound; transplant rejection;

KM graft-versus-host disease; warts; alopecia universalis;

KM autoimmune chronic active hepatitis; inflammatory bowel disease;

KM multiple sclerosis; primary biliary cirrhosis; scleroderma;

KM neurite outgrowth; nerve regrowth; Parkinson's disease;

KM Alzheimer's disease; stroke; traumatic spinal cord; brain injury;

KM peripheral neuropathy; ss.

OS Streptomyces hygroscopicus.

XX location/Qualifiers

XX complement (412..1836)

FT /tag- a

FT /note- "fkbw gene"

FT complement (2020..3579)

FT /tag- b

FT /note- "fkbv gene"

FT 3969..4496

FT /tag- c

FT /note- "fkbz gene"

FT complement (4595..5488)

FT /tag- d

FT /note- "fkb1 gene"

FT 5601..6818

FT /tag- e

FT /note- "fkbe gene"

FT 6808..8052

FT /tag- f

FT /note- "fkbf gene"

FT 8156..8824

FT /tag- g

FT /note- "fkbg gene"

FT complement (9122..9883)

FT /tag- h

FT /note- "fkbg gene"

FT complement (9894..10994)

FT /tag- i

FT /note- "fkbi gene"

FT complement (10987..11247)

FT /tag- j

FT /note- "fkbi gene"

FT complement (11244..12092)

FT /tag- k

FT /note- "fkbi gene"

FT	CDS	complement (12113..13150)	FT		
FT		/note= "1"	FT		
FT		/tag= "fkbp gene"	FT	misc_feature	/tag= "aJ"
FT	CDS	complement (13212..23988)	FT		/note= "KR2"
FT		/tag= "m"	FT		complement (34606..35749)
FT		/note= "fkbc gene"	FT	misc_feature	/tag= "ak"
FT		complement (13452..13662)	FT		complement (28750..28960)
FT	misc_feature	/tag= "n"	FT		/tag= "aI"
FT		/note= "ACP6"	FT	misc_feature	/note= "DH2 (inactive)"
FT		complement (13761..14394)	FT		complement (35749..37144)
FT	misc_feature	/tag= "o"	FT		/tag= "am"
FT		/note= "KR6"	FT	misc_feature	/note= "AT2"
FT		complement (14517..15294)	FT		complement (37145..38296)
FT	misc_feature	/tag= "p"	FT		/tag= "an"
FT		/note= "ER6"	FT	misc_feature	/note= "KS2"
FT		complement (15438..16587)	FT		complement (38371..38581)
FT	misc_feature	/tag= "q"	FT		/tag= "ao"
FT		/note= "dehydratase domain (DH) 6"	FT	misc_feature	/note= "ACP1"
FT		complement (16587..17820)	FT		complement (38677..39307)
FT	misc_feature	/tag= "i"	FT		/tag= "ap"
FT		/note= "acylttransferase domain (AT) 6"	FT	misc_feature	/note= "KR1"
FT		complement (17820..19053)	FT		complement (39442..40609)
FT	misc_feature	/tag= "s"	FT		/tag= "ag"
FT		/note= "KS6"	FT	misc_feature	/note= "DH1"
FT		complement (19116..19326)	FT		complement (40609..41842)
FT	misc_feature	/tag= "t"	FT		/tag= "ar"
FT		/note= "ACP5"	FT	misc_feature	/note= "AT1"
FT		complement (19464..20097)	FT		complement (41842..43093)
FT	misc_feature	/tag= "u"	FT		/tag= "as"
FT		/note= "KR5"	FT	misc_feature	/note= "KS of extender module 1 (KS1)"
FT		complement (20241..21420)	FT		complement (43144..43660)
FT	misc_feature	/tag= "v"	FT		/tag= "at"
FT		/note= "DH5"	FT	misc_feature	/note= "ACP of loading domain"
FT		complement (21420..22653)	FT		complement (43777..44629)
FT	misc_feature	/tag= "w"	FT		/tag= "au"
FT		/note= "AT5"	FT	misc_feature	/note= "ER of loading domain"
FT		complement (22653..23892)	FT		complement (44974..46573)
FT	misc_feature	/tag= "x"	FT		/tag= "av"
FT		/note= "KS5"	FT	CDS	/note= "COA ligase of loading domain"
FT		complement (23992..46573)	FT		46754..47788
FT	CDS	/tag= "y"	FT		/tag= "aw"
FT		/note= "fkbp gene"	FT	CDS	/note= "fkbp gene"
FT		complement (24163..24373)	FT		47785..52272
FT	misc_feature	/tag= "z"	FT		/tag= "ax"
FT		/note= "ACP4"	FT	CDS	/note= "fkbp gene"
FT		complement (24997..26146)	FT		52275..71465
FT	misc_feature	/tag= "aa"	FT		/tag= "ay"
FT		/note= "DH4 (inactive)"	FT	misc_feature	/note= "fkbp gene"
FT		complement (26146..27430)	FT		52362..53576
FT	misc_feature	/tag= "ab"	FT		/tag= "az"
FT		/note= "AT4"	FT	misc_feature	/note= "fkbA gene"
FT		complement (27430..28684)	FT		53577..54716
FT	misc_feature	/tag= "ac"	FT		/tag= "bI"
FT		/note= "KS4"	FT	misc_feature	/note= "AT7"
FT		complement (28750..28960)	FT		54717..55871
FT	misc_feature	/tag= "ad"	FT		/tag= "ba"
FT		/note= "ACP3"	FT	misc_feature	/note= "DH7"
FT		complement (29092..29740)	FT		56019..56819
FT	misc_feature	/tag= "ae"	FT		/tag= "bc"
FT		/note= "KR3"	FT	misc_feature	/note= "ER7"
FT		complement (29869..31018)	FT		56943..57575
FT	misc_feature	/tag= "af"	FT		/tag= "bd"
FT		/note= "DH3 (inactive)"	FT	misc_feature	/note= "KR7"
FT		complement (31018..32185)	FT		57710..57920
FT	misc_feature	/tag= "ag"	FT		/tag= "be"
FT		/note= "AT3"	FT	misc_feature	/note= "ACP7"
FT		complement (32185..33439)	FT		57990..59243
FT	misc_feature	/tag= "ah"	FT		/tag= "bf"
FT		/note= "KS3"	FT	misc_feature	/note= "KS8"
FT		complement (33505..33715)	FT		59244..60398
FT	misc_feature	/tag= "ai"	FT		/tag= "bg"
FT		/note= "ACP2"	FT	misc_feature	/note= "AT8"
FT		complement (33823..34480)	FT		60399..61412
FT	misc_feature		FT		/tag= "bh"
FT			FT		/note= "DH8 (inactive)"

Query Match	6 8%	Score 53.6	DB 21	Length 77536
Best Local Similarity	48.4%	Pred. No. 0.032		
Matches 149	Conservative 0	Mismatches 159	Indels 0	Gaps
Oy	447 gattccggcaacgagctacagcggcgccgagctgatactgctgtaactggagcg	506		
Db	27680 GAAAGTCCAACATCGGCGCACATCTCAGGCGCGCGGGGTGCGCGGCTGTATCAAGATGCT	27621		
Oy	507 cggcgatgatacgcggcgagcgagcgctggccacgcgctggaacttggccggggacacttggga	566		
Db	27620 CCGCCGCAATGCGGCGACGCGACCTCCGCGGACCGTGCACGNGGACACGCGCTCCCGCA	27561		
Oy	567 agtcgagatgacgcgactggaactgtaataactgtctctccgcgcgacagccaccac	626		
Db	27560 GCTTCACATGAGAGCGCGCGCGCGCTCGAATCTTCACCGATGACCGCGCTCGCGGAC	27501		
Oy	627 gtcggtgagcagcgctggaacttgaagaccttcaagcagcgagcggtcccgagctaac	686		
Db	27500 GCACCGCGCGCGCGCGCGCGCGCGCTCGAATCTTCACCGATGACCGCGCTCGCGGAC	27441		
Oy	687 ccgagcgagatgctgatactgcaatcggttggagacgggtctgaactctggaggcggggc	746		
Db	27440 CCTGCTGGAGAGCCACCGCGCGCGGAGCGCGCGCGGAGAGCGCGCTCGAGAGCC	27381		
Oy	747 cggctctgc	754		
Db	27380 CGGTGAGC	27373		
RESULT	6			
AA16766				
ID	AA16766	standard; DNA; 756 BP.		
AC	AA16766;			
XX				
DT	24-FEB-1997	(first entry)		
XX				
DE	Collagen-like polymer sequence D gene 2 encoding unit C2A24C2.			
XX				
KW	collagen; repetitive triad motif; recombinant production;			
KW	photographic; medical; structural; fibre; ds.			
XX				
OS	Synthetic.			
XX				
PN	US5496712-A.			
XX				
PD	05-MAR-1996.			
XX				
PE	06-NOV-1990;	90US-0609716.		
XX				
PR	05-NOV-1992;	92US-0972032.		
PR	06-NOV-1980;	90US-0609716.		
XX	12-NOV-1991;	91US-0791960.		
XX				
PA	(PROT-) PROTEIN POLYMER.			
XX				
PI	Cappello J, Ferrari FA;			
XX				
DR	WPI; 1996-150728/15.			
DR	P-PSDB; AAR93248.			
XX				
PT	Collagen-like polymers comprising repetitive triads - produced in			
XX	unicellular organisms with improved characteristics, useful in, e.g.			
XX	photographic and medical fibres			
XX				
PS	Example 2; Column I7-18; 43pp; English.			

[illegible]

```

FT      /tag- a
FT      /product- "1-deoxy-D-xylose 5-phosphate synthase"
FT      /partial
XX      WO200032792-A2.
XX      08-JUN-2000.
XX      02-DEC-1999. 99WO-US28587.
XX      03-DEC-1998. 98US-0110779.
XX      (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX      Cahoon RE, Tao Y, Williams ME, Coughlan SJ, Weng Z:
XX      MPI: 2000-412338/35.
XX      P-PSDB; AAY97425.
XX      Polynucleotide encoding 1-deoxy-D-xylose 5-phosphate synthase enzyme
XX      useful for producing transgenic plants and for producing antibodies
XX      specific to which is useful for screening cDNA expression libraries -
XX      Claim 2: Page 67; 73pp; English.
XX      The present sequence is a putative coding sequence for the wheat
XX      1-deoxy-D-xylose 5-phosphate synthase enzyme (DXPS). Its protein is
XX      involved in the isoprenoid biosynthesis pathway. The cDNA was identified
XX      by sequencing a number of clones and then comparing their protein
XX      sequences to known proteins: this showed the sequence's similarity to the
XX      Capsicum annum DXPS sequence. The DXPS gene and protein can be used to
XX      create transgenic plants which express the gene at either different
XX      levels or at different stages of development compared to normal, and to
XX      identify herbicides.
XX      Sequence 1404 BP; 277 A; 482 C; 439 G; 206 T; 0 other:
SQ
Query Match 6.3%; Score 49.4; DB 21; Length 1404;
Best Local Similarity 46.9%; Pred. No. 0.18;
Matches 220; Conservative 0; Mismatches 246; Indels 3; Gaps 2;
XX      229 ggaacttcagatcacagcgccgcatcagacaagcgacacagcgccgctatccg 288
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      355 ggcgtcgtcagctcagctcgcgcgtgcactacgtctcaacaccccgacagacaagctc 414
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      289 gccatctactcgtggtcagctggcgccctgcagcagcaatccggatcgccgcgcgc 348
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      415 ctctggagcgtcgccacacgctgtaccgcacacaagattctgacggggcgccgataag 474
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      349 gtgcaggaagctgtcgcagctgcgcagagcttgacgctcacgccgacacgagggcgc 408
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      475 atgcgcagcagtg--cggcagacacacgctgtccgcttcacaaagcgctccag-agc 531
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      409 tggaaatgcgcgtcagacatctgttcagctccggttcacgaattcccgcaacgctaaagc 468
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      532 gagtaacagacgctcggcagccagacacagctccacacacctccgcgcgccttcgagtg 591
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      469 gggcgccgcagctgtatctgtcgtgaacttgaaacggcgcgctgatatcgccggcagc 528
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      592 gccgtctggagggagcctcaagggcgcgaaacaacgctgtgtcggtgattggggagcgg 651
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      529 cgcgttgcacacgttggaactgtccggggccacacttggaagctgtgatatgcgcactggac 588
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      652 gccatgacggcggcgagcggtacgagcgatgaacaacgcggctaccctcgcactcgac 711
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      589 tggaaataacatgcctaccggcgacgacacgccaacacgctcggtgagcgagctggagactg 648
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      712 atgactgtgattctcaacgaacaaacagcaggtgtcgtctgcgcagcgagcagctcagcgg 771
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      649 aagcgcttcacagcagcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 697
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      772 ccggcgccgcgcgcgtggcgcgctcagcgacgacgacccctcagcaagctgtcagct 820

```

```

RESULT 8
AAQ43032
ID      AAQ43032 standard; DNA; 756 BP.
XX
XX      AAQ43032:
XX      07-SEP-1993 (first entry)
XX      Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.
XX      Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
XX      fibre; film; coating; triad sequence; collagen; mammalian; moulding;
XX      hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
XX      Synthetic.
XX      MO9310154-A.
XX      27-MAY-1993.
XX      04-NOV-1992. 92WO-US09485.
XX      12-NOV-1991. 91US-0791960.
XX      (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX      Cappello J, Ferrari FA;
XX      MPI: 1993-182496/22.
XX      P-PSDB; AAR37738.
XX      High mol. wt. collagen-like protein polymers - capable of being
XX      produced in unicellular microorganisms
XX      Disclosure; Page 39; 82pp; English.
XX      The sequences given in AAQ43030-34 encode examples of recombinantly
XX      produced DCP collagen-like polymers (CLPs) which consist of repeated
XX      tripeptide sequences selected from a wide range of GXY sequences,
XX      where X and Y can be any amino acid. These sequences can be cloned
XX      into plasmids and used to transform E. coli to produce the DCP
XX      proteins. DCP peptides comprise repeated units of: A = GAGGPGGP,
XX      B = GSKDPPGP and/or C = GAGGPGPK. These polymers may be used to
XX      raise anti-DCP antibodies in rabbits. These polymers have molecular
XX      weights of >30 kD and are able to form helices due to interchain
XX      linkages. These polymers pref. contain a proportion of tripeptide
XX      triad sequences found in natural collagens, pref. mammalian collagens.
XX      The CLPs impart unique characteristics to materials such as fibres,
XX      membranes, films, coatings, hydrogels, colloid suspensions and moulded
XX      articles.
XX      Sequence 756 BP; 84 A; 288 C; 291 G; 93 T; 0 other:
SQ
Query Match 6.2%; Score 49; DB 14; Length 756;
Best Local Similarity 43.5%; Pred. No. 0.21;
Matches 223; Conservative 0; Mismatches 290; Indels 0; Gaps 0;
XX      123 gtccgaacgttggagcgcgcgcatgtggccggggcgctacacgggtgatacaacagt 182
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      218 gtgcgcttggaccggctgtgtccaccgggtgtctccggacctgacgagccgcgagtgccg 277
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      183 atggggcgcggaagaccggccagtgcatagggtcgcgaacttgaaagggcaactcaacgat 242
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      278 ctggaccggtgttccacccgggtgtcccggaaccttcaagcccgcaaggtgctcctggagc 337
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      243 caacgggcccagatcaacgacgacaaacagctgtgcgcgcctatccggcatctacttcgg 302
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      338 cgcgtgttccacacgggtgtctccgggaccttgcagccgcaggtgctgcctgacgcgctgc 397
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      303 gtccacttggggcgccttgcaggaacattcggattgcggcgcgctgtcaggaagctgtc 362

```

```
Db 398 gttcacccggtgtctccggagaccctcgcccaagctgctccttgacccgctcgttccac 457
OY 363 cgcgcgtgcgcacagctgtgacgtctcaagccgaatcaagacggccgctcgtgataccta 422
Db 458 cgggtgtctcccgagactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 517
OY 423 cgcacatctgtatcagtcctccatcaagatccgcgaacgcgtatcacgcgcgcgcgcgcgcgc 482
Db 518 ctccgcggacctgtgacggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 577
OY 483 gatgatctgctgtgaactggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 542
Db 578 gacctgcagtgcccgccaggtgcgccttgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 637
OY 543 ggaacgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 602
Db 638 caggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 687
OY 603 ctaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 635
Db 698 cgcacagtgcccatgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 730
```

RESULT 9

AAV64548 standard; DNA; 985 BP.

AAV64548;

27-JAN-1999 (first entry)

M. tuberculosis Immunogenic polypeptide XP25 3'-end DNA.

Tuberculosis: immunogenic; soluble; antigen; protective immunity; TB; vaccine; pharmaceutical; infection; diagnosis; ss.

Mycobacterium tuberculosis.

WO9816646-A2.

23-APR-1998.

07-OCT-1997; 97WO-US18293.

13-MAR-1997; 97US-0818112.

11-OCT-1996; 96US-0730510.

(CORI-) CORIXA CORP.

Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;

WPI; 1998-261042/23.

Claim 4: Page 176; 230pp; English.

This sequence encodes an immunogenic portion of a soluble Mycobacterium tuberculosis (MT) antigen which can be used in a method for inducing protective immunity against tuberculosis (TB). This sequence can be formulated into vaccines and/or pharmaceutical compositions for immunising against M. tuberculosis infection or may be used for the diagnosis of tuberculosis.

Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.2%; Score 48.8; DB 19; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.24;

Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```
OY 64 ttcccgatgtgcgaacaaggaaagaaacccggagcctcgagcccgacccgctcggagctg 123
Db 175 ttcgcgcgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 234
OY 124 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 183
Db 235 accggtacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 294
OY 184 tgggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 243
Db 295 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 354
OY 244 acacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 303
Db 355 ggccttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 414
OY 304 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 359
Db 415 atcaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 474
OY 360 gtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 419
Db 475 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 533
OY 420 ctacgcacatctgtgtcagtcctccatcaagatccgcgcgaacgcgcgcgcgcgcgcgcgcgcgc 479
Db 534 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 593
OY 480 gctgatgatctgtgactggaactggaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 539
Db 594 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 653
OY 540 cgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 599
Db 654 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 713
OY 600 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 653
Db 714 cacttgatcgtcgtcagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 767
```

RESULT 10

AAV44439 standard; DNA; 985 BP.

AAV44439;

09-NOV-1998 (first entry)

Mycobacterium tuberculosis antigen XP25 3' DNA.

Tuberculosis: infection; diagnosis; antigen; XP25; ss.

Mycobacterium tuberculosis strain Erdman.

WO9816645-A2.

23-APR-1998.

07-OCT-1997; 97WO-US18214.

13-MAR-1997; 97US-0818111.

11-OCT-1996; 96US-0729622.

(CORI-) CORIXA CORP.

Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;

WPI; 1998-251292/22.

PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis

XX Claim 4; Page 187-188; 250pp; English.

XX This is the 3' region of DNA coding for an antigenic portion of
CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in
CC AAV4438. XP25 DNA was isolated from a M. tuberculosis strain Erdman
CC genomic DNA expression library using sera from patients having
CC extrapulmonary tuberculosis. It bears no similarity to known
CC sequences. The invention relates to methods and compositions for
CC diagnosing tuberculosis. It provides polypeptides (see
CC AAW64291-W64379) comprising an antigenic portion of a M.
CC tuberculosis antigen, or an immunogenic portion of a M.
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers.

XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.2%; Score 48.8; DB 19; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.24;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 64 ttcccgatgagcagaacgaagaaacggagctgagccgagccgagcgtcgagctg 123
DB 175 ttcccgatgagcagaacgaagaaacggagctgagccgagccgagcgtcgagctg 234
OY 124 tgcgacgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 183
DB 235 accggtacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294
OY 184 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 243
DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354
OY 244 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 303
DB 355 ggccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 414
OY 304 tgcgacgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474
OY 360 gtcgacgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 419
DB 475 gccgacgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 533
OY 420 ctacgacatcgtgttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 479
DB 534 cggcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
OY 480 gtcgacatcgtgttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 539
DB 594 ggcgacatcgtgttcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 653
OY 540 cgtggaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 599
DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
OY 600 cgcctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
DB 714 caccgtatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 767

RESULT 11
AA219349
ID AA219349 standard; DNA; 985 BP.

XX AA219349;
AC 05-NOV-1999 (first entry)
DT
XX
DE M. tuberculosis antigen 3' XP25 DNA sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test; ss.

XX Mycobacterium tuberculosis.

XX WO9942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025197.

XX (COR-) CORLXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MD, Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1999-527409/44.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic

PT skin tests and protective or therapeutic vaccines or compositions

XX Claim 4; Page 171-172; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
CC AA219225 are used in the exemplification of the present invention.

XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.2%; Score 48.8; DB 20; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.24;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 64 ttcccgatgagcagaacgaagaaacggagctgagccgagccgagcgtcgagctg 123
DB 175 ttcccgatgagcagaacgaagaaacggagctgagccgagccgagcgtcgagctg 234
OY 124 tgcgacgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 183
DB 235 accggtacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294
OY 184 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 303
DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354
OY 244 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
DB 355 ggccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 414
OY 304 tgcgacgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359
DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474

FT	/product= "ORE5"
FT	complement (5392..6147)
FT	/tag= h
FT	/product= "ORE6"
FT	complement (6152..6156)
FT	/tag= i
FT	complement (6194..7282)
FT	/tag= j
FT	/product= "ORE7"
FT	complement (7280..8133)
FT	/tag= k
FT	/product= "ORE8"
FT	complement (8141..8145)
FT	/tag= l
FT	complement (8254..9318)
FT	/tag= m
FT	/product= "ORE9"
FT	complement (9324..9328)
FT	/tag= n
FT	9575..10504
FT	/tag= o
FT	/product= "ORF10"
FT	9568..9571
FT	/tag= p
FT	10584..11585
FT	/tag= q
FT	/product= "ORF11"
PN	WO200151639-AZ.
XX	
PD	19-JUL-2001.
XX	
PP	12-JAN-2001; 2001WO-US01187.
XX	
PR	12-JAN-2000; 2000US-0175751.
XX	
PA	(SCHE) SCHERING CORP.
XX	
PI	Hosted TJ, Horan AC, Wang TX;
XX	
DR	WPI: 2001-442147/47.
DR	P-PSDB: AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906, AAU04907, AAU04908, AAU04909, AAU04910, AAU04911.
XX	
XX	
PT	New nucleic acid molecules encoding evernimomycin pathway gene
PT	products, useful for improving yields of evernimomycin, to produce new
PT	evernimomycin and as probes to identify homologous sequences -
XX	
PS	Example 1; Fig 12; 109pp; English.
XX	
CC	The sequence encodes 11 proteins comprising enzymes of the
CC	evernimomycin antibiotic biosynthetic pathway. A vector comprising a
CC	M. carponacea evernimomycin biosynthetic pathway resistance gene product
CC	is useful for selecting for a transfected or transformed host cell. An
CC	integrative version of the vector is useful for introducing a
CC	evernimomycin pathway gene (a bottle-neck gene) into an actinomycete of
CC	the genus Micromonospora. The DNA encoding the biosynthetic proteins is
CC	useful for synthesising novel evernimomycin-related compounds, arising
CC	from modifications of the DNA sequence designed to change glycosyl and
CC	modified orsellinic acid groups contained in evernimomycin, for
CC	expressing functional or mutant evernimomycin biosynthetic enzyme for
CC	evaluation, diagnosis and preferably biosynthesis of evernimomycin or
CC	other secondary metabolic products, improving the yield of evernimomycin
CC	and to produce novel evernimomycins and also as a hybridisation probe to
CC	identify homologous sequences. The encoded polypeptides are useful for
CC	combinatorial biosynthesis to generate libraries of orthomycins, e.g.
CC	evernimomycin analogues/homologues and drug discovery. The
CC	DNA encoding the integrase allows for increasing a given gene dosage. The
CC	integrative vector can be used to permanently integrate copies of a
CC	heterologous gene of choice into chromosomes of different hosts and to
CC	integrate genes which increase the yield of known products or to generate
CC	novel products such as hybrid antibiotics or other novel secondary
CC	metabolites. The vector can also be used to integrate antibiotic

[illegible]

XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
 PS
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterium pathogen.
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and
 CC H37Rv (A199682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
 XX
 SO Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
 Query Match 6.1%; Score 48; DB 22; Length 4403765;
 Best Local Similarity 45.6%; Pred. No. 0.62;
 Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;
 QY 151 gccggggggcgccaccgggtgatcaacaagctatggggcgagagcccgccagtcgact 210
 DB 3963782 GTCGGCAACGGGAACTCGGCAACCAACCGGCGAGCGCAATACCGCAACTCCAAAC 3963723
 QY 211 gaagtcgagctggaacagcggaactcacgatacacagcgccgacacagcaagcaac 270
 DB 3963722 GTCGGGCTTCGGGAATACCGGCAACCAACGTCGGCACCGGAAAGCCCGGCAAGCGGCAAC 3963663
 QY 271 aacgtgcccgcctatccggccatctctcgtgtgcacatggggcgccctgcacagcaat 330
 DB 3963662 ATCGCGCGCGGGGAACACCGGCGAGCAAACTGGGCTTCGGCAACACGCGCAATTGGCAAC 3963603
 QY 331 tcgggttcgcgcgcggtgagcagcagcagcagcagcagcagcagcagcagcagcagc 387
 DB 3963602 ATCGGCTTCGGCAACACCGGCGAGCAAACTCGGCTTCGGGCTTCACCGCAACCAACGAG 3963543
 QY 388 acgcgcatacagacagcgccgtggaatgcgcgcctacagacatcgttcagtcgccgtcaag 447
 DB 3963542 GTGGGCAATCGGTGGGCTGAACCTCGGGGAGCGGAACATCGCTTGTCAACTCGGGGCAAC 3963483
 QY 448 aattccggcaacggctacagcgcgcgccgagctgagctgagctgagctgagcaagc 507
 DB 3963482 AACCAACGTCGCTTCTCAACTCCGGCAACGCGCAACTGGGTATGCGCAACTCCAGCGAC 3963423
 QY 508 ggcgtgatgcggcgcgccgcgcgcgccgctgagcagcagctgagcagcagcagcagcagc 567
 DB 3963422 GCGAAATGTCGCAATCGGAACCTCGGGGGCGCACCGTGGGCCCGCTTCGTGGCGGCGCATAC 3963363
 QY 568 gtctgtatgcgcagctgagcagctgaaattacatgcgc 603
 DB 3963362 ACCGCTTCGGGAACCTCGGCTTCGCTCAACACGGGC 3963327
 RESULT 15
 AAT93095
 ID AAT93095 strand: cDNA; 24379 BP.
 XX
 AC AAT93095;
 XX
 DT 11-MAY-1998 (first entry)
 XX
 DE Streptomycetes frenolicin gene cluster.
 XX
 KW Frenolicin; antibiotic; feed additive; anticoccidial;
 KW coccidiostats; efflux pump; butyrate starter synthase;
 KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;
 KW dehydrase; ketoreductase; hydroxylase; Streptomycetes roseofulvus;
 KW ds.
 XX

OS Streptomycetes sp.
 XX
 XX Key
 FT CDS
 FT 636..2948
 /location=Qualifiers
 /tag= a
 /product= 80 kDa non-membrane protein
 /note= "gene A (specifically claimed)"
 2945..3916
 /tag= b
 /product= membrane protein
 /note= "gene B (specifically claimed)"
 4020..4844
 /tag= c
 /product= protein with 6 membrane-spanning domains
 /note= "gene C (specifically claimed)"
 4841..6415
 /tag= d
 /product= ATP-binding component of ABC transporter
 /note= "gene D (specifically claimed)"
 6533..7183
 /tag= e
 /product= unknown non-membrane protein
 /note= "gene E (specifically claimed)"
 7344..8897
 /tag= f
 /product= putative efflux pump
 /note= "gene F (specifically claimed)"
 9164..10012
 /tag= g
 /product= transcription activator
 /note= "gene G (specifically claimed)"
 complement (10105..10621)
 /tag= h
 /product= translationally coupled to gene I
 /note= "gene H (specifically claimed)"
 complement (10618..11628)
 /tag= i
 /product= homologue of fabH
 /note= "gene I (specifically claimed)"
 11809..12066
 /tag= j
 /product= acyl carrier protein
 /note= "gene J (specifically claimed)"
 complement (12154..13209)
 /tag= k
 /product= acyltransferase
 /note= "gene K (specifically claimed)"
 13409..14686
 /tag= l
 /product= PKS ketosynthase subunit
 /note= "gene L (specifically claimed)"
 14767..16047
 /tag= m
 /product= PKS chain length factor
 /note= "gene M (specifically claimed)"
 16120..16371
 /tag= n
 /product= acyl carrier protein used by the PKS
 /note= "gene N (specifically claimed)"
 complement (16453..16935)
 /tag= o
 /product= putative hemiketal dehydrase
 /note= "gene O (specifically claimed)"
 17088..17903
 /tag= p
 /product= ketoreductase related to actIII
 /note= "gene P (specifically claimed)"
 17903..18898
 /tag= q
 /product= cyclase/dehydrase related to act VII
 /note= "gene Q (specifically claimed)"
 18895..19839
 /tag= r
 FT CDS

```

FT FT /product= cyclase/dehydrase related to activ
FT FT /note= "gene R (specifically claimed)"
FT CDS complement (1990..20907)
FT FT /tag= s
FT FT /product= oxidoreductase
FT CDS /note= "gene S (specifically claimed)"
FT complement (20904..22094)
FT FT /tag= t
FT FT /product= quinone-forming hydroxylase
FT FT /note= "gene T (specifically claimed)"
FT CDS complement (22505..22179)
FT FT /*tag= u
FT FT /product= not specified
FT FT /note= "gene U"

EP806480-A2.
XX PD 12-NOV-1997.
XX PF 02-MAY-1997; 97EP-0107329.
XX PR 07-MAY-1996; 96US-0016753.
XX PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX PI Reeves CD, Soliday CL;
XX DR WPI: 1997-538619/50.
XX DR P-PSDB: AAM34199-W34219.
XX PT Streptomyces frenolicin gene cluster - useful for producing
XX PT recombinant frenolicin antibiotics
XX PS Claim 1: Page 40-60; 66pp; English.
XX

This DNA sequence comprises the Streptomyces frenolicin gene
CC cluster containing specifically claimed coding sequences (genes
CC A-U) that respectively encode 21 proteins (see AAM34199-219) involved
CC in frenolicin synthesis. The genes can be divided into 5
CC subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2)
CC genes H, I, J and K encode butyrate starter synthases; (3) genes
CC L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and
CC R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;
CC and (5) genes S and T encode a keto/enoyl reductase and a
CC hydroxylase. Also claimed are vectors, host cells (especially a
CC Streptomyces sp., particularly Streptomyces roseofulvus), and the
CC encoded proteins. Cells transformed using the above sequence can
CC be cultured to produce frenolicins or frenolicin precursors. The
CC precursors can be converted to frenolicins by chemical or other
CC methods. The frenolicins can be oxidised to frenolicin B, an
CC antibiotic used as an anticoccidial agent. The frenolicins can be
CC used as animal feed additives.
XX

SQ Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 other:

```

```

Query Match 6.1%; Score 47.8; DB 18; Length 24379;
Best Local Similarity 47.2%; Pred. No. 0.48;
Matches 145; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

```

```

QY 387 cagcgcatcaagcagcgccgtggaatgcgcctacgaacatctgtcagtcgccgtcac 446
   || || || || || || || || || || || || || || || || || || || ||
DB 17387 caacggcgccgctcggcgagtgacgcgcgaactcagcagctccctgtggaaga 17446
   || || || || || || || || || || || || || || || || || || || ||
QY 447 gaattccggcagcgtcagcgcgccgagcagcagcagcagcagcagcagcagcagc 506
   || || || || || || || || || || || || || || || || || || || ||
DB 17447 cgtcatgcaccacaactcaagaagcgtctcgtgtaccccggaagtgctcacccagcg 17506
   || || || || || || || || || || || || || || || || || || || ||
QY 507 cggcgtgagtcggcgccgagcgtggtccacccgtggaactgcccgggccaacctggga 566
   || || || || || || || || || || || || || || || || || || || ||
DB 17507 cgggagacgcggcgccgagcgtcgtcaacatcgctccacgcggcgcaagcaggg 17566
   || || || || || || || || || || || || || || || || || || || ||
QY 567 agtctgtatgccgactggaattacatcgctacgcgacgagcagccacacac 626
   || || || || || || || || || || || || || || || || || || || ||

```

```

DB 17567 cgtcgtcttcggcgcccccctactcgcgtccaaagcagcgctcgtcgttcaccaaagc 17626
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 627 gtcggtgagcagcgtggaactgaagccttcacgaacgagcgtcgcgcggtacat 686
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17627 cctcgccctcgaaactgcccggagcgcgacataccgtcaacgctcgtcccgctaogt 17686
   || || || || || || || || || || || || || || || || || || || ||
QY 687 ccggcccg 693
   | | | | |
DB 17687 cgagacg 17693

```

Search completed: August 19, 2002, 13:53:47
Job time: 14326 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:28:03 ; Search time 130.83 Seconds
(without alignments)
1475.716 Million cell updates/sec

Title: US-10-003-759-3

Perfect score: 786

Sequence: 1 atgaacgcatcgtcgtcggc.....ttccgtaacggtcagtag 786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

767066

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCMTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	131.8	16.8	1116	4	US-09-104-308-2	Sequence 2, Appl1
2	131.8	16.8	1116	4	US-09-321-981-2	Sequence 2, Appl1
3	105.8	13.5	1716	4	US-09-321-981-4	Sequence 4, Appl1
4	49.8	6.2	736	1	US-08-642-255-50	Sequence 50, Appl1
5	48.8	6.2	985	4	US-09-056-556-182	Sequence 182, Appl1
6	47.6	6.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
7	47.6	6.1	30001	2	US-08-125-468-1	Sequence 1, Appl1
8	47.2	6.0	1908	2	US-08-173-508-1	Sequence 1, Appl1
9	47.2	6.0	1908	2	US-08-265-310-1	Sequence 1, Appl1
10	47.2	6.0	1908	2	US-08-951-742-1	Sequence 1, Appl1
11	46.8	5.9	15872	4	US-09-105-537-1	Sequence 1, Appl1
12	46.2	5.9	1288	1	US-08-440-856A-9	Sequence 9, Appl1
13	46.2	5.9	44377	2	US-08-804-327C-7	Sequence 7, Appl1
14	46.2	5.9	44377	2	US-08-804-198-1	Sequence 1, Appl1
15	45.2	5.8	1291	3	US-08-997-897-1	Sequence 1, Appl1
16	45.2	5.8	1291	3	US-09-156-836B-1	Sequence 1, Appl1
17	45.2	5.8	2852	4	US-09-056-556-203	Sequence 17, Appl1
18	44.6	5.7	1176	2	US-08-387-942C-17	Sequence 17, Appl1
19	44.6	5.7	11220	4	US-09-105-537-32	Sequence 32, Appl1
20	44.6	5.7	12588	2	US-08-387-942C-1	Sequence 1, Appl1
21	44.6	5.7	36778	4	US-09-105-537-5	Sequence 5, Appl1
22	44.6	5.7	38506	3	US-09-320-878-19	Sequence 19, Appl1
23	44.6	5.6	2712	3	US-09-025-691-4	Sequence 4, Appl1
24	43.6	5.5	1017	4	US-08-849-751-1	Sequence 1, Appl1
25	43.6	5.5	1017	4	US-09-478-816-1	Sequence 1, Appl1
26	43.6	5.5	1500	4	US-09-593-711A-10	Sequence 10, Appl1
27	43.4	5.5	1500	4	US-09-593-711A-10	Sequence 10, Appl1

28	43	5.5	1273	4	US-09-319-892-3	Sequence 3, Appl1
29	43	5.5	1333	4	US-09-372-422A-9	Sequence 9, Appl1
30	42.8	5.4	599	1	US-08-584-226-3	Sequence 3, Appl1
31	42.4	5.4	20235	1	US-07-642-734C-3	Sequence 3, Appl1
32	42.4	5.4	20235	3	US-08-439-009A-3	Sequence 3, Appl1
33	42.2	5.4	1158	4	US-09-372-422A-21	Sequence 21, Appl1
34	42.2	5.3	71989	4	US-09-443-501A-2	Sequence 2, Appl1
35	41.8	5.3	432	1	US-08-642-255-48	Sequence 48, Appl1
36	41.6	5.3	1155	2	US-08-387-942C-7	Sequence 7, Appl1
37	41.4	5.3	18994	1	US-08-913-264-2	Sequence 2, Appl1
38	41.4	5.3	18994	2	US-08-459-586-4	Sequence 4, Appl1
39	41.4	5.3	18994	2	US-08-282-696-4	Sequence 4, Appl1
40	41	5.2	1182	4	US-09-385-028-19	Sequence 19, Appl1
41	41	5.2	8051	2	US-08-576-626A-2	Sequence 2, Appl1
42	41	5.2	11604	4	US-09-385-028-13	Sequence 13, Appl1
43	41	5.2	15079	4	US-09-385-028-1	Sequence 1, Appl1
44	41	5.2	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
45	41	5.2	4411529	4	US-09-103-840A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-104-308-2
; Sequence 2, Application US/09104308
; Patent No. 6187577
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: Van Der Kleij, Wilhelmus A.H.
; APPLICANT: Van Solingen, Piet
; APPLICANT: Weyler, Walter
; TITLE OF INVENTION: No. 6187577e1 Cellulase Producing Actinomycetes,
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104, 308
; FILING DATE: 24-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/974, 042
; FILING DATE: 19-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC539
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7555
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-104-308-2

Query Match 16.8%; Score 131.8; DB 4; Length 1116;
Best Local Similarity 50.7%; Pred. No. 9.6e-21;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.M.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 985 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-182

Query Match 6.2%; Score 48.8; DB 4; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.015;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

QY 64 ttcccgatgagcagcaacggaagacgagcctgagccgagccgagcctgagctg 123
DB 175 ttgcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 234
QY 124 tggcagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 183
DB 235 accggtaccgagcagcagcagcagcagcagcagcagcagcagcagcagc 294
QY 184 tggcagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 243
DB 295 gccgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 354
QY 244 accgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 303
DB 355 gccgttaccgagcagcagcagcagcagcagcagcagcagcagcagcagc 414
QY 304 tggcagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 359
DB 415 atmacgggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474
QY 360 gtcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 419
DB 475 gccgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 533
QY 420 ctacgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 479
DB 534 cgcgaacgggagcagcagcagcagcagcagcagcagcagcagcagcagc 593
QY 480 gtcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 539
DB 594 gccgaatggcagcagcagcagcagcagcagcagcagcagcagcagcagc 653

TREATM

QY 540 cgtgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 599
DB 654 cgcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 713
QY 600 cgcctacgagcagcagcagcagcagcagcagcagcagcagcagcagc 653
DB 714 cactgtatcggttaccgagcagcagcagcagcagcagcagcagcagc 767

RESULT 6

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: ERASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 6.1%; Score 48; DB 4; Length 4403765;
Best Local Similarity 45.6%; Pred. No. 0.069;
Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;

QY 151 gccggggggcgtcagcagcagcagcagcagcagcagcagcagcagcagc 210
DB 3963782 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3963723
QY 211 gagcgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 270
DB 3963722 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3963663
QY 271 aacgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 330
DB 3963662 atcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3963603
QY 331 tcgagcgtgagcagcagcagcagcagcagcagcagcagcagcagc 387
DB 3963602 atcgagcagcagcagcagcagcagcagcagcagcagcagcagc 447
QY 388 accgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 447
DB 3963542 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3963483
QY 448 aatcggcagcagcagcagcagcagcagcagcagcagcagcagcagc 507
DB 3963482 aacgaatgagcagcagcagcagcagcagcagcagcagcagcagc 3963423
QY 508 ggcgtatgagcagcagcagcagcagcagcagcagcagcagcagcagc 567
DB 3963422 cgcgaatgagcagcagcagcagcagcagcagcagcagcagcagc 3963363
QY 568 gtcgtatgagcagcagcagcagcagcagcagcagcagcagcagcagc 603
DB 3963362 accggttgcgagcagcagcagcagcagcagcagcagcagcagc 3963327


```
RESULT 7
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy E.
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/125,468
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match 6.1%; Score 47.6; DB 1; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0.044;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

QY 195 gaccgccagtgatgagtcgagtcggaacgaggaacacacacacgagccga 254
DB 25022 GACCCGCAACGCGCTCGCCCTCGCGAGGCGGCGCATGTCGTGTGAGAGCGCGGA 24963
QY 255 tcacgcaacagcaaacagctggcgccctctccgcatctacttccggtgccaactggg 314
DB 24962 GCACGGGATCCGGCGGCGGCGACATCTAGCCGAGGTCACCGGCTACCGCGGCGCTG 24903
QY 315 cgcctcacagcaatctggatcgcgcgcggtgcaagagctgccaactgagcgac 374
DB 24902 CAACCTCTACAGATGACCGGCTCGCGGTGCGACGGAACGAGATGCGCCAGGCTATCCG 24843
QY 375 gaggctgagcgtacacgcgaatcaacgagcgcgctggaatgcgcctacgacactggt 434
DB 24842 GGTCTGCGCTGACCGCGGCGGATGAGACCCGAGGAGCTGCGCTACGTCACCGCGCA 24786
QY 435 caatcccgtaacattccgcaacggtctaacgcgcgcgagctgagctgagctggtc 494
DB 24785 CGGCTGGCGACCAA---GCAGAACGACCGCGACGAGACCGCGCGCTCAAGAGAGACCT 24729
QY 495 gaactggaacgagcggtgtagtcgagcgagcgagcgctgagcaccgtggaactggcgg 554
DB 24728 GGGGACACGCGCTACCGGCTGCGGTGCGGTGAGTCACTGATGATGCGACACTCGCT 24669
```

```
QY 555 ggcacactgggaagtcgtatgcgactggagtcgaaatcatcatcgctaccgagccac 614
DB 24668 GGGCGGGGTCTGCGCCCTGGAGTTGCCCGCACGGGCTGGCGATCGAGACACTCGGTGCT 24609
QY 615 gaccgccaccagtcggtgagcgagctggaactgaagcccttcatacgacgaagcggtgc 674
DB 24608 GCGCGCCAGGCGCCMACCTGCACGTCCCGCATCCCGAGTGGAGACTCGACTACCTGCCGCT 24549
QY 675 ccgagctacatccgagcgagtgatctgcatgctgagtgagagcgaggttcgaactcg 734
DB 24548 GGTGGCGCGGGAACAGCGGCTGCGCACGGTCTCACGCTGCGAGCGGCTTCGCGGCTT 24489
QY 735 ggaaggcgagccgagctc 752
DB 24488 CCAGACGCGCATCTGCT 24471

RESULT 8
US-08-474-933-1/c
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy E.
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match 6.1%; Score 47.6; DB 2; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0.044;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

QY 195 gaccgccagtgatgagtcgagtcggaacgaggaacacacacacgagccga 254
```



```

Db 459 GCGCTGTSGCGSRGRCGCGS---CTSGACGYSWSTGCAACACCGCGGGGCTTG 515
QY 280 gctatccgcatctactctcgtgctgacactgagcgccctgacagcaatctcgatg 339
Db 516 GCGCGGCAACGCGCGCGCGCGCAAGACATCTGTCTTGACGCGSSSAGTTGACGCC 575
QY 340 ccgagcgcgctgcaagagctgtcgaagctgcaagagctgcaagctgcaagctgca 399
Db 576 GTGCTCCGCGTCAAGCGCTGGGCGCGCGCTCGGATGACACCGCGCGCGCGCATG 635
QY 400 agggcgccgtggaatgagccctacatctgttcagctcccgctacagattccggcaac 459
Db 636 ACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 694
QY 460 ggtacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 519
Db 695 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 754
QY 520 ggcggcgagcgcgctgcaagctgcaagctgcaagctgcaagctgcaagctgca 579
Db 755 AACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 814
QY 580 gactggagctggaattacatcgctacacgcaagcgcaagcgcaagcgcaagcgca 639
Db 815 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 874
QY 640 ctgagctgaagcgctctcctcagcagcgcgctgcccgcgagctacatccgcgag 699
Db 875 GACGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 934
QY 700 tctcgaatgagctgagcgagcgctgcaagctgcaagctgcaagctgcaagctg 746
Db 935 AAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 981

RESULT 14
US-08-804-227C-7/C
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

```

```

Query Match 5.9%; Score 46; DB 2; Length 44377;
Best Local Similarity 47.6%; Pred. No. 0.1;
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```

```

QY 70 gatggcgacaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 129
Db 18968 GACGGGGGGGTGGCGCTTGGCGGAGCACCGTGTGATCCCTCGCGCTGGGCGGA 18909
QY 130 cgtggagcgcgcgagctgtgagcgagcgagcgagcgagcgagcgagcgagcg 189
Db 18908 GACGACCGCTGTGTGAGGACCGCGCGGTGTGACACGACGAGTACGGGTTTCTG 18849
QY 190 gcgagaccccgcaatgcatctgagctggaactggaacggaactgcaacagcag 249
Db 18848 GGGGAGTGGCTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18789
QY 250 gccgatcaagacaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 309
Db 18788 GCGGACCTGGGACCGGACTCGGTCAACTCGGCGGACGACGACTGATGACCGCGGTGCGTC 18729
QY 310 tggggcgccctgcagagcaatcgagattgcgagcgagcgagcgagcgagcgag 355
Db 18728 GGGACCGCGCTGCGCCACCGACGAGAGATGGCGCACCTTGTGCGCG 18683

```

```

RESULT 15
US-08-804-198-1/C
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Macintosh 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1

Query Match 5.9%; Score 46; DB 2; Length 4437;
Best Local Similarity 47.6%; Pred. No. 0.1;
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 70 gatggcgacaacggaagaaagacccgagccttgagcccgagccgagcgtcgagctgtgcga 129
DB 18968 GACGGCGGCGCTCGCTTGGCGGAGCACCGTGTGATCCGCTCGCGCTTGCGGCGGA 18909
QY 130 cgtctggagacgacgagatgtgacgggagggcgctacacgggtgatacaaacgatatgggc 189
DB 18908 GAGCAGCGCGCTCGTCAAGCACCAGCGGCGGTGTGACACGACGACGACGCGGTGTGTC 18849
QY 190 gcgagagacgacccagtcgacttgaggtcgaacttgaaacgagcaacttcacgatacacg 249
DB 18848 GGGGAGTGGGTCTGAGCAGCGCGGCGGCGGCGGTGCGGTGACGTCACAGGCGCGTAC 18789
QY 250 gccgatacagacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 309
DB 18788 GCGGACCTCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAC 18729
QY 310 tgggagcgtcgcagagcaattcgggattgcgcgagcgagcgagcgagcgagcgagcgag 355
DB 18728 GGGACCGCGCGCTGCCACGACGAGGAGATGGCGCACCTGTGTCCGG 18683

Search completed: August 19, 2002, 13:45:36
Job time: 13940 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 10:59:49 ; Search time 3821.57 Seconds
(without alignments)
2775.981 Million cell updates/sec

Title: us-10-003-759-3
Perfect score: 786
Sequence: 1 atgaacgtcatgcgtgcgtl.....ttccgtacagtgtagtag 786

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estinu:*
4: em_estinu:*
5: em_estov:*
6: em_estov:*
7: em_estro:*
8: em_estro:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.8	7.7	925	12	CNS0091P
2	56.2	7.2	925	12	CNS0091P
3	55.6	7.1	932	12	CNS00720-
4	54.4	6.9	864	10	BF253875
5	53.8	6.8	964	12	CNS003W6
6	50.6	6.4	700	12	AG060149
7	50.4	6.4	935	12	CNS006XK
8	50.4	6.4	1022	9	BE040131
9	50	6.4	894	12	AG060214
10	50	6.4	1009	12	CNS010EW
11	49.8	6.3	449	10	BE517742
12	49.6	6.3	682	9	AU095988
13	49.2	6.3	534	12	AZ934730
14	49.2	6.3	546	12	AZ935386
15	49.2	6.3	644	9	AU030637
16	48.2	6.1	443	9	AV914669
17	48	6.1	446	10	BM373545

18	48	6.1	559	10	BE405069	BE405069	WHE1215-G
19	48	6.1	588	10	BF255157	BF255157	HVSMF000
20	48	6.1	604	9	AV941745	AV941745	AV941745
21	48	6.1	606	10	BF255408	BF255408	HVSMF000
22	48	6.1	631	9	AL505254	AL505254	AV932837
23	48	6.1	635	9	AV932837	AV932837	AV932837
24	48	6.1	637	9	AV933847	AV933847	AV933847
25	48	6.1	666	10	BF628036	BF628036	HVSMF000
26	48	6.1	719	9	AV935472	AV935472	AV935472
27	48	6.1	830	10	BF254391	BF254391	HVSMF000
28	48	6.1	844	12	CNS0052P	AL056652	Drosophila
29	48	6.1	1101	12	CNS0175Y	AL108860	Drosophila
30	47.8	6.1	307	9	AU070991	AU070991	AU070991
31	47.8	6.1	578	9	AU094712	AU094712	AU094712
32	47.8	6.1	623	9	AU032635	AU032635	AU032635
33	47.6	6.1	517	9	AV933848	AV933848	AV933848
34	47.6	6.1	935	12	CNS006XK	AL066051	Drosophila
35	47.4	6.0	407	10	BE475922	BE475922	946049606
36	47	6.0	657	10	BI171319	BI171319	103101981
37	47	6.0	677	9	AL505169	AL505169	AL505169
38	46.8	6.0	269	10	BE636900	BE636900	WHE1805-1
39	46.8	6.0	641	9	AU093820	AU093820	AU093820
40	46.8	6.0	669	9	AU096005	AU096005	AU096005
41	46.8	6.0	674	9	AU055758	AU055758	AU055758
42	46.6	5.9	691	9	AL504955	AL504955	AL504955
43	46.6	5.9	839	12	CNS004NB	AL054280	Drosophila
44	46.4	5.9	453	10	BI779131	BI779131	EBRC01_SO
45	46.4	5.9	557	10	BF618612	BF618612	HVSMF000

ALIGNMENTS

RESULT 1
CNS0091P/c
LOCUS
DEFINITION
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TERT end of BAC #
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION
AL053013
VERSION
AL053013.1 GI:4934461
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 925)

AUTHORS
TITLE
JOURNAL

COMMENT

Genoscope.
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Osoegawa and
Aaron Mammeter in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source

1. 925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"

QY	562	1gggaagcctctgtatgcgcgcagctgagcagctggaattatcatcgtctaccgagcagccgac	621
DB	375	TTTCGAGTTCTGGACCCACGAGTGGGAGAAAGCAGCGCACCTGCTCCAACTGGACCAACGAC	434
QY	622	accagctcgctgagcagctgagcactgagcgttcacatcagcagcagcgcgctgcgccgcgc	681
DB	435	GGCTACTTCGGACAGCGGCTGCTGGCTTCAAGGCCGCCCAACACTCATCATCTATCTCGGC	494
QY	682	tacatcgcgcgcggaagtgtta 701	
DB	495	GACGCCGGGATGTCGCCGTA 514	
RESULT	5		
CNS003WG/C			
LOCUS			
DEFINITION	CNS003WG	964 bp	DNA linear GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence TET3 end of BAC #		
VERSION	BACR09E09 of RPci-98 library from Drosophila melanogaster (fruit		
KEYWORDS	fly), genomic survey sequence.		
SOURCE	AL065254		
ORGANISM	AL065254.1 GI:4942606		
REFERENCE	GSS.		
AUTHORS	fruit fly.		
TITLE	Drosophila melanogaster		
JOURNAL	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Euryptera; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
COMMENT	1 (bases 1 to 964)		
	Genoscope.		
	Direct Submission		
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage		
	BP 191 91006 EVRI cedex - FRANCE (E-mail : seget@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the Berkeley Drosophila Genome Project (BDGP).		
	The BDGP is constructing a physical map of the Drosophila		
	melanogaster genome using these BACs. For further information		
	please see http://www.fruitfly.org The BDGP Drosophila		
	melanogaster BAC library was prepared by Kazutoyo Osoegawa and		
	Aaron Mammeter in Pieter de Jong's laboratory in the Department of		
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
	NY. The library is named RPci-98 and was constructed by partial		
	ECORI digestion of Drosophila DNA provided by the BDGP from the		
	isogenic strain Y2: cn bw ap, the same strain used for the BDGP's		
	PI and EST libraries. A more detailed description of the library		
	and how to order individual BAC clones, the entire library, or		
	filters for hybridization from the BACPAC Resource Center can be		
	found at http://BACPAC.med.buffalo.edu/drosophila_bac.htm .		
FEATURES			
source	1..964		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone_lib="RPci-98"		
	/clone="BACR09E09"		
	/note="end : TET3"		
BASE COUNT	234 a 160 c 60 g 105 t 405 others		
ORIGIN			
Query Match	6.8% Score 53.8; DB 12; Length 964;		
Best Local Similarity	21.4% Pred. NO. 2.2;		
Matches	91; Conservative 154; Mismatches 180; Indels 1; Gaps 1;		
QY	337	ttgcgcgcgcgcgttcagcagctgctgcgaagctgacagcagcgttgacgttcacgcacgcgcgc	396
DB	954	TTTCGSCGCGCYKTTTCSCGSGSTBTSCTSTSSCGSCCBGSGSGSBBSTSGKGSST	895
QY	397	acgaaggcgcgcgtgaaagcgcgcctacacacatctggttcagctcccgctacgaattccgcgc	456
DB	894	SGSTSTVBGCSVSGSSSTSGSSSSBGSNCCGSGCTCBSSCGSGCCGCMKCBTMTTMS	835
QY	457	aacgcctacagcgcgcgcgcgcagcagctgcatctgctgaactggaacgcgcgcgtgcatg	516

Db	834	AMMSSSGSGSGSSGGSVSRTRTKRRSSSGBSGCKSGCGRAGAGGGGGGGRG	775
Qy	517	ccggcgcggaagccgctgacacgctggaactgycgcgggcccactgggaactcgtgat	576
Db	774	SSKSTSGSGSGGGRTRTTTKGSSSSGSGGTTCGSGCGGAASGSGRGVAGBBGTTCT	715
Qy	577	ggcgcgctggaactggaattacatcgtccctccggcgcagcagccaccacgctgaac	636
Db	714	GSTTSSKSGSKTTTSSGKXNMTSSSYCHSTSGASRGTTBSSAAAAGNS-SAASMKC	656
Qy	637	gaagctggaactgaagcccttcacgcagcgcgctgcgccgcgctacatccgcgcgag	696
Db	655	SASVYGVSCCMCASSSCCSSSCMWCGCGCTKCCGSSSVSRGSSRTGCGVCGSSG	596
Qy	697	tggctctgcacgctgtagagcgcgctcgaaactctgggaaggcgcgctgtgcga	756
Db	555	SBTTTSSGGSVAAAACSCAGCAGKGGSSVGGCCGTTMSSRDCAAGACHRVSSGTYA	536
Qy	757	agcgcc 762	
Db	535	VABST 530	
RESULT	6		
LOCUS	AC060149	700 bp	DNA
DEFINITION	Pan troglodytes DNA, clone: PTB-047122.R, genomic survey sequence.		
ACCESSION	AC060149		
KEYWORDS	AG060149.1 GI:16611379		
SOURCE	GSS (genome survey sequence), Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC library clone:PTB-047122.R.		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	BAC end sequences of Library PTB		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 700)		
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tokoki,Y., Watanabe,H. and Sakaki,Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Shuhiro-Chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbges@sc.riken.go.jp, URL:http://npg.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Red process and may have higher chance of clone tracking errors.		
	PRIMERS		
	Sequencing: MJ3rev		
	LIBRARY		
	Vector : pKS145		
	R.Site 1 : Sact		
	R.Site 2 : Sact		
FEATURES	Location/Qualifiers		
Source	1..700		
	/organism="Pan troglodytes"		
	/db_xref="taxon:9598"		
	/clone="PTB-047122.R"		
	/sex="male"		
	/cell_type="lymphoblast"		
	/clone_11b="PTB Chimpanzee Male BAC library"		
BASE COUNT	30 a 291 c 334 g 15 t 30 others		
ORIGIN			
Query Match	6.4%	Score 50.6;	DB 12; Length 700;
Best Local Similarity	42.7%;	Prod. No. 8.2;	
Matches 243;	Conservative 0;	Mismatches 323;	Indels 3; Gaps 1;

[illegible]

P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	SOURCE	location/Qualifiers
	1. .935	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/clone_lib="RPCI.98"	
	/clone="BACRI4N09"	
	/note="end : 77"	
BASE COUNT	257 a 170 c 162 g	96 t 250 others
ORIGIN		

Query Match	6.4%	Score 50.4	DB 12	Length 935
Best Local	Similarly	29.0%	Pred. No. 9	
Matches 123	Conservative	110	Mismatches 184	Indels 7
				Gaps 2

QY	131	gcttggagacgcgcgtgatttgcccggggggcgtacacggtgattcaacaacgtattggcg	190
Db	934	gscggggggsssssgssgcgsgcsgsssgccscscgscscscscscscscscscscscs	875
QY	191	cygaagaccgcgccagttgattgagtgctgcgactgtgaacaggggaacttcaatcacacg	250
Db	874	scscgscgcgcscscgscs	817
QY	251	ccgattcaagaaagcgaacaacgtgacgcgtctccgcgcattcttccgatttcgtgtgcact	310
Db	816	ccggccgcgcscgc	757
QY	311	ggggcgccctgcacgcgaacttgcggatttcgcgcgcgcgcgttgcgaagactgtccgacgtgc	370
Db	756	sgscgscsssgsgscgcgcgcgcgcscscscsgcscgcgcscscgcgcgcgcgcgcgcgcgc	702
QY	371	gcacgagactgtgacgctacgcgcgcgtacgcagacggtgcgctgtgaattgcgcctacgacatc	430
Db	701	scgsgsgcgsgc	642
QY	431	gattcagttccgcgtacgaatttcgcgcgaacggtctacgacgcgcgcgcgcgcgcgcgcgcgc	490
Db	641	ssggcccccgc	562
QY	491	gactgaactgtgaacgycgcgtgtgatttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	550
Db	581	gsssssgasgc	522
QY	551	ccgcg 554	
Db	521	scss 518	

RESULT	8			
LOCUS	BE040131/c			
DEFINITION	BE040131	1022 bp	mRNA	linear
ACCESSION	OD1020208	OD	<i>Oryza sativa</i>	cDNA 5', similar to gata-binding transcription factors, mRNA sequence.
VERSION	BE040131			
KEYWORDS	BE040131.1	GI:8335147		
SOURCE	EST.			
ORGANISM	<i>Oryza sativa</i> .			
REFERENCE	<i>Oryza sativa</i>			
AUTHORS	Enkaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; <i>Oryza</i> .			
	1 (bases 1 to 1022)			
	Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferreira,H., Kawasaki,S., McColough,A., Michalowski,C.B., Palactio,C., Scara,G., Wheeler,M. and Zepeda,G.R.			
TITLE	Functional Genomics of Plant Stress Tolerance			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Michalowski,C.B.			
	University of Arizona			

Query Match 100.0%; Score 732; DB 6; Length 786;
 Best Local Similarity 100.0%; Pred. No. 4,7e-95;
 Matches 732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgaactgctcttcccgatgacgaacgaagaacgcctgaagcccgagcg 60
 |||
 DB 52 TCGGACTGGCTCTTCCCGATGGCGAACGGAAGAACCGGACCTGAGCCGAGCCG 111
 |||
 QY 61 accgtcgaactgtgcgaacgtcggagcgcgcgcatgtgcccggggggccacccgggtg 120
 |||
 DB 112 ACCGTGAGCTGTGCGGACCTGGAGCGCGCGATGTGGCCGGGGGGCCCTACCGGGTG 171
 |||
 QY 121 atcaacaacgctatgaggcgcgagaccccgacgtgcatltagg tgcgaactggaacagggc 180
 |||
 DB 172 ATCAACAACGTATGGGGCGCGGAGACCGCCGACATGATGAGTGGGACTGGAACGGGC 231
 |||
 QY 181 aacttcagatcacaacggcgccgatcaagaacgcgaacacgctgacccctatccggcc 240
 |||
 DB 232 AACTTCAGATCACACGGGCGCATCACGACGGAACGGAACGCTGGCCGCTATCCGGCC 291
 |||
 QY 241 atctactcgggtgacacgtggcgccctgacgagcaattccgagattgcccggcggtg 300
 |||
 DB 292 ATCTACTGGGGTGGCCACTGGGGCGCTGCACAGCAATTCCGGGATTGCCGGCGGTG 351
 |||
 QY 301 caggagcgtctcgaacgtgcgacgagctggaacgtcaacgcga tcaacgagcgcgcttg 360
 |||
 DB 352 CAGGAGCTGTCCGACGTGGCGCACGAGCTGAGCCTCACGCCGATCACAGAGCGCGCTGG 411
 |||
 QY 361 aatgcgcgcacgaacacatctgttcaatgccgttaacgaattccggcaacgctacagggc 420
 |||
 DB 412 AATGCCGCTACGACATCTGGTTAGTCCCGTACAGCAATTCGGCAACGGCTACAGGGGC 471
 |||
 QY 421 gggcgacgactga tga tctgctgaacttgaacagcgcgctg tga tgcggcgcgacggc 480
 |||
 DB 472 GGGCGGAGCTGATGATCTGGCTGMACTGAAGAGCGCGCTGATGCCGGCGGCGAGCGCC 531
 |||
 QY 481 gtggcaccgttgaacgtgcggggcgcaacttgggaag tctggtatgcgaactggagacttg 540
 |||
 DB 532 GTGGCCACCTGGGAAGTGGCGCGGGGCCACCTGGGAAGTGTGATATGCCGACTGGAGCTGG 591
 |||
 QY 541 aattacatgccttaacggcgacgaacgcccacacgctcggtgagcgagctcgagacgtgaag 600
 |||
 DB 592 AATTACATCTCCCTAACCGGCGCACGAGCCACACAGTGGTGAAGGAGCTGAAG 651
 |||
 QY 601 gcttcacgaacgacgagcggtgcgcgcgcggtacatccggcgagatggtatctgcgaatg 660
 |||
 DB 652 GCCTTCATCGACACGCGGCTGCCCGCGGCTACATCCGGCGGAGTGATCTGCAATGGC 711
 |||
 QY 661 gtggagacgggcttcgaacttgggaagcgcggtctgcgaagcgcgca tttttcc 720
 |||
 DB 712 GTGGAGACGGGCTTCAAACTGGGAGGGGCGGCGGTCTGCAAGCGCGCATTTTTC 771
 |||
 QY 721 gtaacggtgcaag 732
 |||
 DB 772 GTAACGCTGCAG 783
 |||

RESULT 2

AX339680 1877 bp DNA linear PAT 10-JUN-2002
 LOCUS AX339680
 DEFINITION Sequence 1 from Patent WO0196382.
 ACCESSION AX339680
 VERSION AX339680.1 GI:18135682
 KEYWORDS
 SOURCE Rhodothermus marinus.
 ORGANISM Rhodothermus marinus
 Bacteria; CFb group; Rhodothermus group; Rhodothermus.
 REFERENCE 1 (sites)
 AUTHORS Wicher, K.B., Holst, O.P., Hachem, M.Y., Karlsson, E.M. and
 Hreggvidsson, G.O.
 TITLE Thermotable cellulase
 JOURNAL Patent: WO 0196382-A 1 20-DEC-2001;

Prokaria ehf. (15)
 FEATURES
 source Location/Qualifiers
 1..1877
 /organism="Rhodothermus marinus"
 /db_xref="taxon:29549"
 710..1495
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAD20684.1"
 /db_xref="GI:18135683"

BASE COUNT 331 a 582 c 632 g 332 t

Query Match 100.0%; Score 732; DB 6; Length 1877;
 Best Local Similarity 100.0%; Pred. No. 3,7e-95;
 Matches 732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgaactgctcttcccgatgacgaacgaagaacgcctgaagcccgagcg 60
 |||
 DB 761 TCGGACTGGCTCTTCCCGATGGCGAACGGAAGAACCGGACCTGAGCCGAGCCG 820
 |||
 QY 61 accgtcgaactgtgcgaacgtcggagcgcgcgcatgtgcccggggggccacccgggtg 120
 |||
 DB 821 ACCGTGAGCTGTGCGGACCTGGAGCGCGGATGTGGCCGGGGGGCCCTACCGGGTG 880
 |||
 QY 121 atcaacaacgctatgaggcgcgagaccccgacgtgcatltagg tgcgaactggaacagggc 180
 |||
 DB 881 ATCAACAACGTATGGGGCGCGGAGACCGCCGACATGATGAGTGGGACTGGAACGGGC 940
 |||
 QY 181 aacttcagatcacaacggcgccgatcaagaacgcgaacacgctgacccctatccggcc 240
 |||
 DB 941 AACTTCAGATCACACGGGCGCATCACGACGGAACGGAACGCTGGCCGCTATCCGGCC 1000
 |||
 QY 241 atctactcgggtgacacgtggcgccctgacgagcaattccgagattgcccggcggtg 300
 |||
 DB 1001 ATCTACTGGGGTGGCCACTGGGGCGCTGCACAGCAATTCCGGGATTGCCGGCGGTG 1060
 |||
 QY 301 caggagcgtctcgaacgtgcgacgagctggaacgtcaacgcga tcaacgagcgcgcttg 360
 |||
 DB 1061 CAGGAGCTGTCCGACGTGGCGCACGAGCTGAGCCTCACGCCGATCACAGAGCGCGCTGG 1120
 |||
 QY 361 aatgcgccttaacgaacatctgttcaatgccgttaacgaattccggcaacgctacagggc 420
 |||
 DB 1121 AATGCCGCTACGACATCTGGTTAGTCCCGTACAGCAATTCGGCAACGGCTACAGCGGC 1180
 |||
 QY 421 gggcgacgactga tga tctgctgaacttgaacagcgcggtg tga tgcggcgcgacggc 480
 |||
 DB 1181 GGGCGGAGCTGATGATCTGGCTGMACTGAAGAGCGCGCTGATGCCGGCGGAGCGCC 1240
 |||
 QY 481 gtggcaccgttgaacgtgcggggcgcaacttgggaag tctggtatgcgaactggagacttg 540
 |||
 DB 1241 GTGGCCACCTGGGAAGTGGCGCGGGGCCACCTGGGAAGTGTGATATGCCGACTGGAGCTGG 1300
 |||
 QY 541 aattacatgccttaacggcgacgaacgcccacacgctcggtgagcgagctcgagacgtgaag 600
 |||
 DB 1301 AATTACATCTCCCTAACCGGCGCACGAGCCACACAGTGGTGAAGGAGCTGAAG 1360
 |||
 QY 601 gcttcacgaacgacgagcggtgcgcgcgcggtacatccggcgagatggtatctgcgaatg 660
 |||
 DB 1361 GCCTTCATCGACACGCGGCTGCCCGCGGCTACATCCGGCGGAGTGATCTGCAATGGC 1420
 |||
 QY 661 gtggagacgggcttcgaacttgggaagcgcggtctgcgaagcgcgca tttttcc 720
 |||
 DB 1421 GTGGAGACGGGCTTCAAACTGGGAGGGGCGGCGGTCTGCAAGCGCGCATTTTTC 1480
 |||
 QY 721 gtaacggtgcaag 732
 |||

|||||
Db 1481 GTAACGGTCAG 1492

RESULT 3
LOCUS RMU72637 1874 bp DNA linear BCT 11-MAY-1999
DEFINITION Rhodothermus marinus cellulase (celA) gene, complete cds.
ACCESSION U72637
VERSION 072637.1 GI:2304960
KEYWORDS
SOURCE Rhodothermus marinus.
ORGANISM Rhodothermus marinus.
REFERENCE 1 (bases 1 to 1874)
AUTHORS Hallorsdottir, S., Thorolfsson, E.T., Spilliaert, R.,
Hallorsdottir, S.H., Palsson, A.,
Johansson, M., Thorbjarnardottir, S.H., Holst, O. and Eggertsson, G.
Heggvasson, G.O., Kristjansson, J.K., Holst, O. and Eggertsson, G.
Cloning, sequencing and overexpression of a Rhodothermus marinus
gene encoding a thermostable cellulase of glycosyl hydrolase family
12

TITLE
JOURNAL Appl. Microbiol. Biotechnol. 49 (3), 277-284 (1998)
MEDLINE 98242392
REFERENCE 2 (bases 1 to 1874)
AUTHORS Hallorsdottir, S.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1996) Department of Molecular Biology,
Institution of Biology, Grensaevegur 12, Reykjavik 108, Iceland

FEATURES
source
1..1874
/organism="Rhodothermus marinus"
/strain="T1378"
/db_xref="taxon:29549"
710..1493
/gene="celA"
710..1492
/gene="celA"
/EC_number="3.2.1.4"
/note="belongs to family H of cellulase catalytic domains;
endo-1,4-beta-glucanase"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="AAB65594.1"
/db_xref="GI:2304961"
/translation="MNVMRAVLVLSLLFLGCDLPPDGNKEPEPEPEPTVELGCR
WDARDAAGRGYRVINNVGAEFTACQIEVLEGTENFTTRADHNGNNVAAYPAITYGC
HMARPARIRPCARAGAVRAHELDVPTITGDMNAAYDIWSPVNTSGSGAGEL
MTMLNNGCYMPCGSRVATYELGATMEVYVYADMDNRYIYRRTTPTTSVELDKAF
IDVAVARGITRPFMTLHRAVETGELMDEGGAGLRTADPSVTVQ"

BASE COUNT 331 a 580 c 631 g 332 t
ORIGIN

Query Match 94.6% Score 692.8; DB 1: Length 1874;
Best Local Similarity 99.3%; Pred. No. 1.4e-89;
Matches 727; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 1 tgcgactgctcttcccgatgacgacaaagaaacccgagccttgaccgacgcg 60
|||||
Db 761 TGCACAGCGGCTTCCCGATGGCGACACGAAAGACCGGACCTGACCCGACCG 820

QY 61 acgcgtgagctgtgcgacgcttgagacgcgcgacgtgacgcggggggcgctaccgggtg 120
|||||
Db 821 ACCCTCAGCGCTGTCCGACGCTGGGACGCGCGCATGTGGCCGGGGGCGCTACCGGGTG 880

QY 121 atcaacaacgtatgaggcgagaccccgacgtgacgttgagtcgagcttgaaacgggc 180
|||||
Db 881 ATCAACAACGATGGGCGGAGACCGCCAGTGCATGTGAGTGGACTGGAACGGGC 940

QY 181 aactcagatcacagcgacgacacagacaaagcgaacacgagcgccctaccgccc 240
|||||
Db 941 AACTTCAGATCAGACGCGGCGATCAGACAAAGCAACGTCGCGCTATCCGGCC 1000

QY 241 atctactcgggtgacactgaggcgccctgacagacgaatctgggattgcggcggtg 300
|||||
Db 1001 ATCTACTTCGGGTGCCACT-GGGCGCCTGCACGACGAAATCGGGATTG-CGGCGCGCTG 1058

QY 301 caggagctgtcgcagcgtgacgacgagcttgagacgtcaacgcgcgtacagagggccgctg 360
|||||
Db 1059 CAGAGCTGTCCACGCGGCGACGAGCTGGACG-TCACGCCGATCACGACGGGCGCCTGG 1117

QY 361 aatgcgcacacacactgttccagtcgccgacgaattccggcaacggtcacaagcgcc 420
|||||
Db 1118 AATGCCCTTACGACATCTGTTCAGTCCGCTACGAAATTCGGCAACGGCTACACGGC 1177

QY 421 ggcgcgagctgatactgctgacgtgacgtgaaagcgcggtgacgcggcgacgcgcg 480
|||||
Db 1178 GCGCGCAGCTGATGATCTGGTGAACGTGAACGGCGCGCTGATGCGGCGGCGACCGC 1237

QY 481 gtggcgcgctggaactgagcggcgccgacacttggaagctcgtatgctgcgactgagctg 540
|||||
Db 1238 GTGGCCACCGTGAACGTGGCGGGCCACCTGGGAAGTGTGTATGCCGACTGGACTGG 1297

QY 541 aattacatgcctacacgacgacacacacacacacacacacacacacacacacacacac 600
|||||
Db 1298 AATTACATCGCCTTACCGCGCGACGACGCCACACACGTCGAGTGAACGCTGGAAG 1357

QY 601 gcttcacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 660
|||||
Db 1358 GCCTTCATCAGACGACGCGGCTGCGCCGCGGCTACATCCGCGGAGTGTATGTCATGCG 1417

QY 661 gtggagagcggtctgacactctggagggcgggcggtctgcgaagcgccgatttcc 720
|||||
Db 1418 GTGAGAGCGGGCTTCGAACCTGGAGGGCGGGCGCGCTGTGGAAAGCGCATTTTTC 1477

QY 721 gtaacggtgcag 732
|||||
Db 1478 GTAACGGTCAG 1489

RESULT 4
LOCUS AF233376 1116 bp DNA linear BCT 02-AUG-2000
DEFINITION Streptomyces sp. 11A68 cellulase 12A (cel12a) gene, complete cds.
ACCESSION AF233376
VERSION AF233376.1 GI:9651812
KEYWORDS
SOURCE Streptomyces sp. 11A68.
ORGANISM Streptomyces sp. 11A68.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Actinomycetales; Streptomyces; Actinobacteria; Actinobacteridae;
Bacteria; Firmicutes; Streptomyces; Streptomycetaceae; Streptomyces.
van Solingen, P., Melijer, D., van der Kleij, W.A.H., Barnett, C.C.,
Bolle, R., Power, S.D. and Jones, B.E.
Cloning and expression of an endocellulase gene from a novel
Streptomyces isolated from an East African soda lake

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1116)
AUTHORS van Solingen, P., Melijer, D., van der Kleij, W.A.H., Barnett, C.C.,
Bolle, R., Power, S.D. and Jones, B.E.
Direct Submission
TITLE Submitted (10-FEB-2000) Microbial & Molecular Screening, Genencor
JOURNAL International B.V., Archimedesweg 30, 2333 CN Leiden, The
Netherlands

FEATURES
source
1..1116
/organism="Streptomyces sp. 11A68"
/strain="11A68"
/db_xref="taxon:133452"
1..1116
/gene="cel12a"
1..93
misc_feature
/gene="cel12a"
/note="Region: signal sequence"
1..1116
CDS

Oy	647	ggtatctcgtacgtcggttgagacggcgcttcgaaactcttggaggcgggccgcgtctcgcaa	706
Dd	671	ggtacctttaccagacattccaggcgccgtttcgaaacctgggaaggcgccacacgtctggccg	730
Oy	707	gocgccattttccgttaacggtgca	731
Dd	731	tgaactcgtctctctcccgcgggtgaa	755
RESULT	6		
LOCUS	ARI68360	1116 bp	DNA linear PAT 17-DEC-2001
DEFINITION	Sequence 2 from patent US 6287839.		
ACCESSION	ARI68360		
VERSION	ARI68360.1	GI:17904235	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1116)		
AUTHORS	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Meyler,W.		
TITLE	Cellulase producing actinomycetes, cellulase produced therefrom and method of producing same		
JOURNAL	Patent: US 6287839-A 2 11-SEP-2001;		
FEATURES	Location/Qualifiers		
source	1..1116		
BASE COUNT	191 a 432 c 340 g 153 t		
ORIGIN	/organism="unknown"		
Query Match	18.0%; Score 131.8; DB 6; Length 1116;		
Best Local Similarity	50.7%; Pred. No. 6.9e-10;		
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps			1:
Oy	47	cttagcccgagccgaacgctcgagctgtgcgcgtctggagacgcgcgcatgtgcgcggg	106
Dd	77	cggctcccgccacgagccgaacacgacgatctggacgctacggccacacacagatccagg	136
Oy	107	ggcgctcacgggttatcaacaagtatggggcgggagagacggccccgaftgcatltggatcg	166
Dd	137	accggttacgtgtgcagaaacacccgttggggaccacgacgcccacccaagtcatacattga	196
Oy	167	gacttgaaaacgggcaacttcacatcacacacgggcccgaatcacagaacagcgcaaacgltg	226
Dd	197	ccggcaacgggtttcgagatcaaccaacggccgaacggttcggatgccgaaccaagsgcccca	256
Oy	227	ccgcatcacgggcaactactactctgggtgcgcaatggggcgccctgcagagcaatlcgggat	286
Dd	257	agtcctatccctccggtctacgacggcgtccactacggaacactgcgcgcccccgacgacgc	316
Oy	287	tgcggcgcgcggtgtaaaggactcgttcgcagcgtygcagacgagctcgacgctcaagccgatca	346
Dd	317	tgcccatccggatcagctcgcgatccggacacggccccaagttctctcctacccgctacaccg	376
Oy	347	cgaacggcgcgcttgtaatgcgcgcctacgaacatctgttcaagtcgcgtcaagaaattccgga	406
Dd	377	gcaacggcgctctacacgacccgggtacgacatctgg-----ctggaccggacaccccgca	430
Oy	407	acgagctacacggcgcgcgccgagcgtgatgatcttgctgaacttgaaacggcgcggtgatgc	466
Dd	431	ccaacggcggtgaacccggacccggatcagatctggttcaacggcggtcgcccggtccagc	490
Oy	467	cggggcggaagcgcgctgtagcaccgtggaacatgcccggggcccaaccctgggaagtctgtatg	526
Dd	491	ccattcgggttcgcgcggcttcggcagcgccacgctcggcgccgacgcttggaaggttgaccgc	550
Oy	527	ccgactggagcttggaattacatcgcctacacggcgacagcgcccaaccacgtctgctgacg	586
Dd	551	gcacgaacgggtttacgaacgacgtgactcttcttccggccctcctccgcatcacacactgga	610
Oy	587	agctggaccgtgaaggccttcacgaacgacggtccgcgcggtacatatccggccggagt	646

Db	611	GCTTGACGTACAGGACTTTCGTACACAGGCCGCTAGCAGGCGCTGGCCACACCCGGACT	670
Qy	647	ggtatctgcgcatgctggtgagacgaggtcttcgaactctggaagcgagcggtctgcgaa	706
Db	671	GGTACTTCACACGATCCAGGGGGGCTTCGAGCCGTGGAGAGGGCGGCACCGTCTGGCGG	730
Qy	707	gcgcgcaatttctcgtaacggtgca	731
Db	731	TGAACTCGTTCCTCCGCCGGTGAA	755
RESULT	7		
LOCUS	ARI68362	1716 bp	DNA
DEFINITION	Sequence 4 from patent US 6287839.		linear
ACCESSION	ARI68362		
VERSION	ARI68362.1	GI:17904238	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1716)		
AUTHORS	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.		
TITLE	Cellulase producing actinomycetes, cellulase produced therefrom and method of producing same		
JOURNAL	Patent: US 6287839-A 11-SEP-2001;		
FEATURES	location/Qualifiers		
source	1..1716		
BASE COUNT	307 a 611 c 559 g	239 t	
ORIGIN			
Query Match	14.5%;	Score 105.8;	DB 6; Length 1716;
Best Local Similarity	53.5%;	Pred. No. 3e-06;	
Matches	269; Conservative	0; Mismatches	227; Indels 7; Gaps 2;
Qy	230	ccatacgcggccatactcgggtggtcgaactgagcgccctgcagagaaatccggatgac	289
Db	711	CCTATCCCTCGGTCTACGACGCGCTCCCACTACGCAACTCGCGCCGCGACAGACCTGC	770
Qy	230	cgcgcgcgtgcagagagctgtccgaagtcgcgcagtcgcaagcttgaaagctcaagccgatcaga	349
Db	771	CCATCGGATCAGCTCGATCGGACACGCGCCACGACGATGTCCTTACCGCTACACCGGCA	830
Qy	350	cgggcgcgttgaaatgcgcgtcaagcaatctcgtgtcaagcccgtaacgaatccgcgaacg	409
Db	831	ACGGGCTTACACCGCGCGTACGACATCTGGCTG-----GACCCGACACCCCGCACCA	884
Qy	410	gctacagcgcgcgcgagctgagctgagctgagctgaaagcgcgcgctgagtcgag	469
Db	885	ACGGGGTGAACCGGACCCAGATCATGATCTGGTTAACCGGGTGGCCCCGGTCCAGGCCCA	944
Qy	470	gcgcgacgcgcgtggtccacgcgttgaaactgcgcgagcgagccacactggaagctctglatgcg	529
Db	945	TGCGTTCGCGGTCGCGACGGCCACGTCGCGCGCGCCGACGCTGGAGGCTGTGGACCGGCA	1004
Qy	530	actggtgactggaatlaactcgtcctacgcgcgcagcagcccaaccacgctcgtgtgagcagc	589
Db	1005	GCAACGGTTCGAACGACGTGATCTCTTCGCGCCCTCCGCGATACAGACCTGGAGCT	1064
Qy	590	tggaactgaaagcccttaacgcg-cgaagcgctgcgcgcgagctataatccgcgcgagtcg	648
Db	1065	TGCGAGCTGAAGGACTTTCGTGACCAAGGCGGTGACGACCGCGCTGGCCACCCCGGACTGG	1124
Qy	649	tatctgcatacgctggaagacggtcttcgaactctggaagcgagcgcggtctcgtgcgaagc	708
Db	1125	TACCTTCACACGATCCAGGCGGGCTTCCAGCCGTGGAGGAGGGGACACCGTCTGGCCGTG	1184
Qy	709	gcgcgaatttctcgtaacggtgca	731
Db	1185	AACCTGTTCTCCTCCGCGGTGAA	1207

LOCUS	SHU51222	3527 bp	DNA	linear	BCT 23-JUN-1997
DEFINITION	Streptomyces halstedii beta-1,4-endoglucanase (celA2) gene, complete cds.				
ACCESSION	U51222				
VERSION	U51222.1	GI:2209259			
KEYWORDS					
SOURCE	Streptomyces halstedii.				
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
REFERENCE	1 (bases 1 to 3527) Garda, A.L., Fernandez-Abalos, J.M., Sanchez, P., Ruiz-Arribas, A. and Santamaria, R.I.				
AUTHORS	Two genes encoding an endoglucanase and a cellulose-binding protein are clustered and co-regulated by a TPA codon in Streptomyces halstedii JMB				
TITLE	Biochem. J. 324 (Pt 2), 403-411 (1997)				
JOURNAL	97307849				
MEMLINE	2 (bases 1 to 3527)				
REFERENCE	Santamaria-Sanchez, R.I.				
AUTHORS	Direct Submission				
TITLE	Submitted (13-MAR-1996) Ramon I. Santamaria-Sanchez, Instituto de Microbiologia Bioquimica, Consejo Superior de Investigaciones Cientificas / Universidad de Salamanca, Avda. del Campo Charro, 9/n, Salamanca, Salamanca 37007, Spain				
FEATURES	Location/Qualifiers				
SOURCE	1..3527				
gene	/organism="Streptomyces halstedii"				
CDS	/strain="JMB"				
	/db_xref="taxon:1944"				
	595..1728				
	/gene="celA2"				
	595..1728				
	/gene="celA2"				
	/EC_number="3.2.1.4"				
	/note="beta-1,4-endoglucanase"				
	/codon_start=1				
	/transl_table=1				
	/product="Cel2"				
	/protein_id="AAC54549.1"				
	/db_xref="GI:2209260"				
	/translation="MRALPGPRPRSLIGALILATVAAVLAAPAAHADLTLYCEO YGSTTIGRIVYQNNRNGASAPQCVRTTDSGFRYQADGAVPTNGAKSPSPNGCH WLNKVPGLPIPTGSAQAGTASVAGFRTMQRKNSGNDVISEFVAVASWSEDFVFR NLNARQANNMNYLTYSVOAGFEPWQNGAVLNSFSTVNLGTFGTGSPGAPGPYA CEVYATNWSNDSGFTADVTVAANTGSTPEVDNMRSLFTLPSORVYNAWNAVTSPTGAV TATGLASAOVAAAGASOTFFGQGYSTFEPKPSFSLNGARCA"				
	1945..3039				
	/note="cellulose binding protein"				
	/codon_start=1				
	/transl_table=1				
	/product="P40"				
	/protein_id="AAC5430.1"				
	/db_xref="GI:2209261"				
	/translation="MARGKRLVSLIVATVTLGGILTLFGQAGNAHGVTYTPGSR TYLCMDAKTSTGSLDPTNPKRKALESQSNLLYNNFAVLDSNAGRGAGVYDPKRL CSADRSPLYNTGYNARSMPRTHLTAGRTIOVKSNNMAHAPGSFVYLSKGYSPS TELCMDLELTETYNPQPGSPCTDGGHYWNLDLPSGRSGDAVMF IOWVRSDSOEN FEFSGDVYFPGNGEYVTVGIRGSGSTPRPTPTPPDPTPTDPSGCMAYVRYATNWS GGPGSVENWNNHSTRTADGNAVKYKTPGCAKAKVSSWNGALTTGSDGAVYRSLDYNRS IPPGGSYTFEGFTATSTGNNLPLVSGIGCVNP"				
	BASE COUNT 623 a 1270 c 1106 g 528 t				
ORIGIN					
Query Match	13.4% Score 97.8; DB 1; Length 3527;				
Best Local Similarity	48.5% Pident No. 3.3e-05;				
Matches 304; Conservative	0; Mismatches 317; Indels 6; Gaps 1;				

OY	105	gggagcgctaacccgggtgatcaacaagaatctggggcgcgggaacccgcagttgatatggagt	164
Db	747	GGGCGCGTATGTGTCACAGAACCAACCGCTGGGGCCGGAATGCCCCCCAATGGTGTAACCBC	806
OY	165	cggacttggaaaaggcgaaacttcacgatcacacaggggcgcatctcagacaacggaacaacgt	224
Db	807	TACGCAGAAGCGGCTTCCGGGTGATCCACAGGCCACGCGCGCGTTCCCAAGAAAGTGCCCC	866
OY	225	ggcgcgctataccggccaatacttccgttgccattcggggcgccctcyaagagcaattcggg	284
Db	867	GAAGTCCTATCCTCCCTGGCTTTCMACGGCGTGCATCACCAACAATCTCTCCCGGGAGCGAA	926
OY	285	attcgcgcgcgcgcfygcgaggaigtglccgacgttgcgcgaacgaigtgagctcaacgcgat	344
Db	927	CCTCCCGGCGCAGGTGAGCGCGTATGCCAGGCCCCCACACACATCTCTTACGCGCTTTGT	986
OY	345	caagacggggcgccgttgaatgcgcgcctaacaacalcgtgtlcaagtlcccgttaogaatlccgg	404
Db	987	CGGCAAGCGCGGTGTMCACACGCGTGTGATGACATCTGGCGTGGAGCCCAACCAAGAGAA	1046
OY	405	caaagcgctaacagcggcgcgccgagctgatctctgcttgaacttgaacgtgaacgycggtgat	464
Db	1047	CGGGGTGAAC-----GTACCGAGATCATGTATCTGGCGTCAACAAGGTGCGGCCATCCA	1100
OY	465	ggcggggcgcgacgcgcgfygccccagctgaaacttgcgcgggggcacacttgaaagtctgta	524
Db	1101	GCCCATTCGGCTTGCGACGCCCGCACCGCCCTCCGTCGGCGGGCCACTGCGAGTGTGAG	1160
OY	525	tgcgcacttggagcttgaattacatctgcctaccgcgcgacagccccaccaacgaactgtag	584
Db	1161	GGGAGCAACAGGCTCCACAGACGATCATCTCTTCCGCGCCCTTCGGCGCTGCGCAGCTG	1220
OY	585	cgagcttgcacctgaagcgcccttatctgaacgacgcgttcgcccgcgctacatccgcggga	644
Db	1221	GAGCTTCAGACGTCAAGGACTTCTCCGGAACACATCGCCCGCGCATGGCGCAACA	1280
OY	645	gtggtatctgcacatgcygagagcgggtctcaactctcgyagggcggggcgctctgcg	704
Db	1281	CTGGTATCTCACCAATGTTCACGACAGAGATTCAGACGTCGACGAAGGTGCCGGACTGCC	1340
OY	705	aagcgcgatlttccgtaaacygtca	731
Db	1341	GGTGAACCTCTTCTCTCAACCGTAGA	1367
RESULT	9		
SREGLS			
LOCUS		1470 bp	DNA linear BCT 16-DEC-1994
DEFINITION	S.rochei egls gene.		
ACCESSION	X73953		
VERSION	X73953.1 GI:393391		
KEYWORDS	cellulase; egls gene; endoglucanase.		
SOURCE	Streptomyces rochei.		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
	Bacteria; Fimicutes; Streptomycinae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 1470)		
AUTHORS	Mastromei,G.		
JOURNAL	Direct Submission		
	Submitted (07-JUL-1993) G. Mastromei, Universita di Firenze, Dipt		
	di Biologia Animale e Genetica, Via Romana, 17/19, 50125 Firenze,		
	ITALY		
REFERENCE	2 (bases 1 to 1470)		
AUTHORS	Pelito,B., Hanhart,F., Irdani,T., Iqbal,M., McCarthy,A.J. and		
	Mastromei,G.		
TITLE	Characterization and sequence analysis of a Streptomyces rochei A2		
JOURNAL	endoglucanase-encoding gene		
MEDLINE	Gene 148 (1), 119-124 (1994)		
FEATURES	95011642		
	Location/Qualifiers		
	1..1470		
	/organism="Streptomyces rochei"		

```
/strain="A2"
/db_xref="taxon:1928"
/clone="E.coli pcsf1"
90..103
repeat_unit
/rpt_type=INVERTED
112..125
repeat_unit
/rpt_type=INVERTED
152..157
RBS
166..1314
/gene="egls"
166..1314
CDS
166..1314
/gene="egls"
/EC_number="3.2.1.4"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="CAA52139.1"
/db_xref="GI:393392"
/db_xref="SPTREMBL:O59963"
/translation="MRLRHHRLRAVSAALLTALAALATATAPADDTTICEE
FGSVYIOGRVYVQNNRMCTSAOCVTATDSGFRVYQADGSYPTNGAPRSYSEVNGCH
YTNCSPGTALPARISGISAPSSISYGVNDAVYNAVDIMLDPTRDGNVREITMI
WENRYGOIOPISGOWGTA SVAGRMEVMSGGCTDYLSPVAPMSMSNEDPMFVR
ATVARGLAGNDMYLTSLIOAFEPWONGGLAVNSFSTVNTGSGNFGDPNGDPTGT
PACTVSTATNVMPGGFTANVTYVNNGSAPVDGMLATFLPSGQSVHAHMAVSPPSS
GAVATGPAESRLIAAGSOSFQGAYSGFAPAPARQLNCTACSTV"
stem_loop
1321..1340
stem_loop
1351..1403
stem_loop
1419..1453
BASE COUNT 211 a 579 c 475 g 205 t
ORIGIN
```

```
Query Match 13.1%; Score 96.2; DB 1; Length 1470;
Best Local Similarity 47.3%; Pred. No. 7.2e-05;
Matches 327; Conservative 0; Mismatches 358; Indels 6; Gaps 1;
```

```
QY 41 cggagcctgagcccgagccgagccgtgagctgtgagagcgtggaagcgcgagatgtg 100
DB 254 CCGCGACCGCCCGCCGACGAGCCGACACACGATCTCGAGAGTTCGGCTCGACCGTCA 313
QY 101 cgggagggcgctacggcggtgatacaagcgtatggggcgcgagagcccgagtgatcg 160
DB 314 TCCAGGGCGCTTACGTCGTCAGAACACCGCTGGGGCACACCGCCACCGACGTCGTA 373
QY 161 aggtcgagctggaagcggcaacttcacgatacacagggcgatcaagaacggcaaca 220
DB 374 CCGGACCGACCTCCGGCTTCGGGTCACGACGAGCCGACGCTCGGTCGGACCAAGCGG 433
QY 221 acgtgagcgctatccggccatctacttcggtgtgccaactggggcgctgcaagcaat 280
DB 434 CCGCGAAGTCGACCGCTGCTTCACGCGCTGTCACACAGAACTGTCGCGCGGCA 493
QY 281 cgggattcgcgcgcgcgcgagagctgtccgagctgagcgagctgagagctgagcgc 340
DB 494 CCGCGCTCCCGCGCGGATCAGCGGATCTCTCCGCGCCACACGACATCTGTACGGCT 553
QY 341 cgaatcagcagggcgctggaatgcccctacgacatactgtgtcaagtcgccgtcaagca 400
DB 554 TCGTCAGAACGCGCTGTACAAAGCGCTCGTACGACATCTGGCT-----GAAACCGA 607
QY 401 cgggcaagcgctacagcgcgcgcgagcgtgatactgtgctgaactggaacgagcgagc 460
DB 608 CCGGACCGACGCGGGTGAACCGGACCGAGATCATGTATCTGTTCAACCGGGTGGGCA 667
QY 461 tgatgagcgagcgagcgagcggtgagcagctggaactggccggcgagcagctggagct 520
DB 668 TCCAGCGCATTCGGCTGCGAGCTGCGACGAGCTTCCTGGCGCGGCGCCACCTGGAGAGT 727
QY 521 ggtatgcgagctggaacttatactgcctacacggcgagcagcagcaccacgctcg 580
DB 728 GGTCCGCGGGAACGACCAACGAGCTGCTGTGCTGCTGCTCGCTCGCGATGAGCA 787
```

```
QY 581 tgagcagctggagcctgaagccttcacgacgagcggtgcgcccgtacatccggc 640
DB 788 GGTGAGCTTCCAGCTGATGAGCTTCGCCGCGACCTGCGCCGCGGCGTGGCGGCA 847
QY 641 cggagtgatctgctgagcggtgagcagcggttcgaacttggagggcgagcgcg 700
DB 848 ACGACTGATACCTGACGAGACATCCAGGCGGTTTCAGCCCTGGCAGAGAGGTGGGTC 907
QY 701 tgcgaagcgcgagcttttcgtaacggtgca 731
DB 908 TCGCGTCAACTCTCTCTCTGCGACCGGTGA 938

RESULT 10
AF130408 1311 bp mRNA linear BCT 14-APR-1999
LOCUS
DEFINITION Streptomyces viridosporus strain T7A cellulase (celSI) mRNA,
complete cds.
ACCESSION AF130408
VERSION AF130408.1 GI:4583444
KEYWORDS
SOURCE
ORGANISM Streptomyces viridosporus.
Streptomyces viridosporus.
Bacteria; Filicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
1 (bases 1 to 1311)
REFERENCE Ramchandran,S. and Crawford,D.L.
AUTHORS Characterization and sequence analysis of two genes involved in
TITLE cellulase degradation in Streptomyces viridosporus T7A, and its
expression in Escherichia coli
JOURNAL Unpublished
2 (bases 1 to 1311)
AUTHORS Ramchandran,S. and Crawford,D.L.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1999) Department of Microbiology, Molecular
Biology and Biochemistry, University of Idaho, LSN 125, Gibb Hall,
Moscow, ID 83844-3052, USA
FEATURES
source
1..1311
/organism="Streptomyces viridosporus"
/strain="T7A"
/db_xref="taxon:67581"
1..1311
/gene="celSI"
repeat_region
82..97
/function="putative regulatory region"
144..149
RBS
144..149
/gene="celSI"
/feature="putative"
157..1311
/gene="celSI"
/feature="celSI; family H member"
/EC_number="3.2.1.4"
/transl_table=11
/product="cellulase"
/protein_id="AAD25090.1"
/db_xref="GI:4583445"
/translation="MRLRAPDPAPCAVMSPPCSPLSPSRMRPRRRAADDTTICEP
YGTTLIOGRVYVQNNRMCTSAOCVTATDSGFRVYQADGSYPTNGAPRSYSEVNGCH
YTNCSPGTALPARISGISAPSSISYGVNDAVYNAVDIMLDPTRDGNVREITMI
WENRYGOIOPISGOWGTA SVAGRMEVMSGGCTDYLSPVAPMSMSNEDPMFVR
ATVARGLAGNDMYLTSLIOAFEPWONGGLAVNSFSTVNTGSGNFGDPNGDPTGT
PACTVSTATNVMPGGFTANVTYVNNGSAPVDGMLATFLPSGQSVHAHMAVSPPSS
GAVATGPAESRLIAAGSOSFQGAYSGFAPAPARQLNCTACSTV"
misc_feature
166..168
/gene="celSI"
/feature="rare tta codon"
/function="putative regulatory function"
BASE COUNT 216 a 494 c 414 g 186 t 1 others
ORIGIN
```


Query Match	12.2%	Score: 89	DB: 1	Length: 1311
Best Local Similarity	52.1%	Pred. No. 0.00078		
Matches: 225	Conservative	0	Mismatches: 201	Indels: 6
				Gaps: 1
QY 270	caacgacaaattccggaatttcgacgagccgctgacgaagctcttcgacgctgacgacgagctg	329		
DB 504	caagcgacattcttcgacgagccgacgagctgacgaagctcttcgacgagctgacgacgagctg	563		
QY 330	gaacgctcaacgacgacgacgagccgctgacgaagctcttcgacgacgacgacgacgacgacg	389		
DB 564	cttcgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	617		
QY 390	cgacgacaaattccggaatttcgacgagccgctgacgaagctcttcgacgacgacgacgacg	449		
DB 618	ggacacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	677		
QY 450	gaacgagcggcgctgacgacgagccgctgacgaagctcttcgacgacgacgacgacgacgacg	509		
DB 678	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	737		
QY 510	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	569		
DB 738	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	797		
QY 570	caacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	629		
DB 798	ggacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	857		
QY 630	caacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	689		
DB 858	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	917		
QY 690	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	701		
DB 918	cgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg	929		
RESULT 11				
LOCUS SLU04629	1565 bp	DNA	linear	BCR 02-Oct-1997
DEFINITION Streptomyces lividans 66 cellulase B (celB) gene, complete cds.				
ACCESSION U04629				
VERSION U04629.1	GI:2462717			
KEYWORDS				
SOURCE ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

Query Match	Best Local Similarity	Score	DB 1:	Length
Matches	298; Conservative	0; Mismatches	324; Indels	6; Gaps
QY 105	ggggcgctacccggygatcaacaagatgagggcgcgagagaccgccaagtgcattgagtc	12.0%; 47.5%;	DB 1:	1565;
Db 489	ggggcagatcagctgctcagacacacacggcggggctcaccgcccccacgctcagcggc			
QY 165	cggacttgaaacggcgcaacttcagatcacacagggcgcatccgcgaacagcgacaacgct			
Db 549	caccgacacacggccttccggctcagcagccacgacgctcgccacccgacccacggcgcc			
QY 225	gacgcctatccgcgcatacttcggtgtgcacatcgagggcgccctcagagaaatctggg			
Db 609	gaagctcgtacccgctggcttcaacggcgctccacatcagacactgtttacacggcgacga			
QY 285	attgcgcgagcgctgtcagagagctgtccgaagtgcgcaagcttgaaagctcaagcgcat			
Db 669	ccctcccgctccgctcagacaccgctcccgccgcccgcctcaccacatctgtagcgcttgc			
QY 345	cacgacgggacggcgtgaaatgcgcctcagacatctcggttcagttccgctcagaaatccgg			
Db 729	cgacggccgctgtacacacgctcgtacgacataatgg-----ctggacccgacggcccg			
QY 405	caacgctacagcggcgcgccgagctgatactgtgctgaaatcgaaacggcgcgctgcat			
Db 783	caccgacggcggtgaaacacagaccgaatcatgtctggttcaacaggggtggctccgatcca			
QY 465	gacggggggcgagcgcggtggccacgctggaacttgcgcgggggccacatcggaaatctgta			
Db 843	gcccatcgccctacggctgacggccgctccgctccgcccggacctgggaagcttggag			
QY 525	tgccgctgagagctggaatatacatcgcttaccggcgacgacgacccacacagctcggtgag			
Db 903	cgccggcaacggcctcgacagacgctgctgtgcttggggacacgctggcgcatcagggcgtg			
QY 585	cgaagcttgagacctgaagcgcttcatcgagaagcggtctgcggcggtctatactcgccgga			
Db 963	gaagcttgcacgctcaatgagatttgcctggcgccgacccgttcgacacggaataacga			
QY 645	gtgtgatctgcacgtcggtgagaaagggcttcgaaactctcggaagggcgggccggtctcg			
Db 1023	cttgctactcgaacgacgcttcagggcggttgcagccgcttgcgacacggcccggaactggc			
QY 705	aagcgccgaatttccgtaacggtgcag			

Db 1083 CGTGAACCTCTCTCTCCTCCACGCTCGAG 1110

RESULT 12

LOCUS SCG11A 41782 bp DNA linear BCT 21-MAR-2000

VERSION AL133210

ACCESSION AL133210.1 GI:6468421

KEYWORDS acyl-CoA dehydrogenase; aldehyde dehydrogenase; amidase; dihydroxy-acid dehydratase; DNA polymerase III beta chain; export protein; gntR-family transcriptional regulator; lacI-family transcriptional regulator; marX-family transcriptional regulator; oxidoreductase; reductase; regulatory protein; secreted cellulase B precursor; secreted cellulose binding protein; tetr-family transcriptional regulator; transcriptional regulator; xylose repressor; xylose kinase.

SOURCE Streptomyces coelicolor A3(2).

ORGANISM Streptomyces coelicolor A3(2).

REFERENCE 1 (bases 1 to 41782) Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

AUTHORS Redenbach,M., Kieser,H.M., Denaplatte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.

TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)

REFERENCE 2 (bases 1 to 41782) 97000351

AUTHORS Oliver,K. and Harris,D.

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 41782) Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.

AUTHORS Direct Submission

JOURNAL Submitted (23-NOV-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

COMMENT Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)

CDS are numbered using the following system eg SC787.01c. SC (S. coelicolor), 787 (cosmid name), 01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/juncgi-bin/frameplot.pl>.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc, ttg or (atc)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid G11A lies on the AseI-G genomic restriction fragment.

Location/Qualifiers

1..41782

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid F37"

1..1299

/gene="xy1B"

1..101

/gene="xy1B"

/note="nominal overlap with Streptomyces coelicolor cosmid 28tG11"

<1..1299

/gene="xy1B"

/note="SCG11A.01, xy1B, xylose kinase, len: >432 aa; highly similar to SW:XYLB_STRRU (EMBL:M73789) Streptomyces rubiginosus xylose kinase (EC 2.7.1.17) (xylose kinase) xy1B, 481 aa; fasta scores: opt: 2553 z-score: 2795.9 E(): 0; 89.2% identity in 427 aa overlap. Contains match to pfam entry PF00370 FGXY, FGXY family of carbohydrate kinases and two prosite matches to entries PS00933 FGXY family of carbohydrate kinases signature 2"

/codon_start=1

/trans_table=11

/product="xylose kinase"

/protein_id="CAB61582.1"

/db_xref="GI:6468422"

/translation="DPRQWMDALGELASQCGEAREAAVSVGGQOHLVTLDARGEPVPALLMDVRSAPQARLIDELGAKAWEATGVSFVTWAMLTETPEAPARAVKAVRLPHDYLTERLTGEGTIDRGDVGSTGMAWSTGEVDEILARALDPAALPRVVRPGVACVGRDGHLPESKGLVACGDNMAAALGIGRCGVVMSLGSCTGYAVSRRPADPPTGYAGFADARQDMLPLACTLNLCTLVANDRVASLGLDREAVPETDTTLPLLDEKTPNLPHSSGLLGHLDHTAGQLQAAVDGAVSHLGLDLVLDDADPSPA PLLLIGGARGATMOOVTRLSGRVQIPEARELVALGAAQAAGLITGDEDAVAARRNNTAAGPYDAVERDEATLNRITGVLSDAAPLERDASR"

76..1122

/gene="xy1B"

/note="Pfam match to entry PF00370 FGXY, FGXY family of carbohydrate kinases, score 514.30, E-value 8.7e-151"

250..288

/gene="xy1B"

/note="PS00933 FGXY family of carbohydrate kinases signature 1"

874..936

/gene="xy1B"

/note="PS00445 FGXY family of carbohydrate kinases signature 2"

1519..1523

1525..2733

/gene="SCG11A.02"

1525..2733

/gene="SCG11A.02"

/note="SCG11A.02, possible xylose repressor, len: 402 aa; similar to SW:XYLR_BACSU (EMBL:M27248) Bacillus subtilis xylose repressor xy1R, 384 aa; fasta scores: opt: 612 z-score: 642.4 E(): 2.1e-28; 29.4% identity in 381 aa overlap. Contains match to Pfam entry PF00480 ROK, ROK family and a possible helix-turn-helix motif at residues 43..64 (+3.15 SD)"

/codon_start=1

/trans_table=11

/product="putative xylose repressor"

/protein_id="CAB61583.1"

/db_xref="GI:6468423"

/translation="MSAPPEAOPARPGRALPDTOGMRRLISRYMHTVSAEGLPSRAVVASRIGLTRAASVTLVDLIRSGLLELGEPRGVRPGSALAVSQGPGAGAEVGDHLAVCAVDLRGVRARAVRGSRSRPEVLEQITGVROYVSGAETEGILMPAGLAAYPGIVARDGRVYVRAPIUDMHDADLGLALPADLPPTYDNEANREAGALAEMLGDGTRPDLHVASRIGTGAAYVFDGRLLRGTGAGELGHVPVDPKRCAGGRCLEQVGEKAVYLAAAGVPEGEDRVGLIARAAAGDDVRALAEACTAIGLITGVNVLDP EGVVLGALAGLAPMLPLSLRDELARRTGAPACPAVAVSRLGPGQLGAHSVYAVLDDPGAVAERA"

1822..2373

/gene="SCG11A.02"

	gene	/note="Pfam match to entry PF00480 ROK, ROK family, score 131.1:80, E-value 4e-38"
	CDS	3041..3646 /gene="SCG11A_03" 3041..3646 /gene="SCG11A_03" /note="SCG11A_03, probable amidase, len: 201 aa: similar to SW:AMPD_ECOLI (EMBL:X15237) Escherichia coli amydro-N-acetylmuramyl-L-tripeptide amidase Ampd, 183aa: fasta scores: opt: 211 z-score: 267.1 E(): 1.7e-07; 39.5% identity in 114 aa overlap and to Streptomyces coelicolor SC211.21c; fasta scores: opt: 674 z-score: 751.6 E(): 0; 50.0% identity in 202 aa overlap. Contains match to Pfam entry PF01510 Amidase_2, N-acetylmuramoyl-L-alanine amidase"
	misc_feature	/codon_start=1 /transl_table=1 /product="putative amidase" /protein_id="CAB61584.1" /db_xref="GI:6468424" /translation="MERBRTAPVPRRLKGAALAVPYPTLSCTRAANQAVDP SAENLPASTNSYSSRRPAPVPDPPVVVHYTQETADLSTFPDEKOVASHVNRSS DGHAAQCRTEDLIAHAGNDWYNTRSGIEHWGRDFEFTNMYBQSARLTACTTA AYGPKDRTHLIAHEPVSDHTDPGFMDTRYIRLVNFA" 3191..3637 /gene="SCG11A_03" /note="Pfam match to entry PF01510 Amidase_2, N-acetylmuramoyl-L-alanine amidase, score 89.80, E-value 5.7e-23" 3811..3815 3821..5080 /gene="SCG11A_04" 3821..5080 /gene="SCG11A_04" /note="SCG11A_04, possible transcriptional regulator, len: aa: similar to TR:O31551 (EMBL:Z99108) Bacillus subtilis transcriptional regulator Acor, 605 aa; fasta scores: opt: 393 z-score: 446.5 E(): 1.7e-17; 34.7% identity in 199 aa overlap" /codon_start=1 /transl_table=1 /product="putative transcriptional regulator" /protein_id="CAB61585.1" /db_xref="GI:6468425" /translation="MTDAWLALPEGADVPERARALRRAHETFEAGTVRRPVAHVAD SWRSRVAGVPGDTASVELMDGLGYRAHPHSRVPLVRELLGTFFADGEILLNV CDAGRILNWGVGPATRRRARRRMNFVGARGESAVGTNAVAVRGVQVAAEH FIRROPWTCAPAAYVDHPIRGVLYGADITGGDLAHPSLGFOVARAESOLALL TPEESAABAELTALGREBALISADGVRILSRSHSEIVYLIAHPGLIGDELCLAL YEDTEPVPTLRAEIARLRICTLTGGRGLASRPYRLTMVESPTSVERRAGVNTAA AAYGCLIPGSOAQAVGRLRRRLADGRLAILACGPDLDLANHAHMGEDDLDMFRA LAAYRPATISSRLALESELAFNPW" 5205..5209 5218..6741 /gene="thca" 5218..6741 /gene="thca" /note="SCG11A_05, thca, aldehyde dehydrogenase, len:534 aa; highly similar to SW:THCA_RHOSN (EMBL:U017129)
Query Match		12.0%; Score 87.6; DB 1; Length 41782;
Best Local Similarity		47.5%; Pred. No. 0.00047;
Matches 298:	Conservative	0; Mismatches 324; Indels 6; Gaps 1;
OY	105	gggagcgctaccggtgatcaacaacgltatggggcggaagacgcscagtcataggct 164
DB	27828	GGCGAGGTACGTCGCCAGAACAACCGCGTGCGCCACCAGCACCGCATGCTCACGC 27887
OY	165	cggacttgtaaaagccggaacttcagaatcaaacgagcgagccaatacgaacgaact 224
DB	27888	CACGCACACACGGCTTCCGGGTACACCCAGCCGACGCTCGGCACCCGACCAAGGGGCC 27947
OY	225	ggcgccatatccggccatatcttgcgggtgcgaactggggcgcttcagagcaattcgg 284

DB 27948 GAAGTGTACCCGCTGGTGCTTTCAACGGCTGCCACTACACGAACCTTTACCAGCGGCACCGC 28007
QY 285 atlgcgcggcgccgctgcaaggaagtctgcacgaatgtgcacgaagcttgaagcttcacgcat 344
DB 28008 COTCCCGTCGCGGCTCGAACACCGCTTCGCGGGCGCGCTCGACGATCTGTACGCTTGC 28067
QY 345 caeagcaggcgacttgaaatgccgcctcacagaacatctggtttcaagtcgcgcacgaattccg 404
DB 28068 GCACGCGCCCGCTTACAACAGCGCTCTCTACGACATAAG-----CTGGACCCGACGCGCCG 28121
QY 405 caeagcgtaccagcggcgagcgccgaagctgatctcgttgtacgtgaagctgaagcgcggtgat 464
DB 28122 CACCACGCGGGGGAACACAGACCGAGATATCATCTGCTTCAACAGGCTGGGTCGATCCA 28181
QY 465 gccggcgagcagcgcgcttgcacccgtgaaacttgccggcgccacacttggaagtcgtta 524
DB 28182 GCCCATCCGGCTCACCGGAGGCGACGGCCGCTCGCGCGCGGACCTGGAGAGTGTGGAG 28241
QY 525 tgcgcagcagtgaaatgaaatcatactgctaaccgscagcgagcccacacaagtcgtgtgag 584
DB 28242 CGCGGGCAACGGCTTGAACAGAGTCTCTGTTCTGCGACCGTGTGGGATACAGCGCTG 28301
QY 585 cgaagcttgacctgaagcgcttcatacgaagcagcggtctgcgccccgagctacatccgcgcca 644
DB 28302 GAGCTTCGACGTCATGAGACTTTCGTCGCGGCGCAGCGCTGCGCGGAGACTCGCGGAACGA 28361
QY 645 gtgtgattctgcatacggtgtgagagacgagcttcgaaacttcggagaggcggtgtctgtcg 704
DB 28362 CTGGTAACCTTGACGAGCGCTTCAAGCGGGGTGTGAGCGCTTGCAAGAAGCGCGCGGACTGCGC 28421
QY 705 aagcgcgatlttcgtaacggtcag 722
DB 28422 CGTGAACCTCTTCTCTCTCCACCGCTGAG 28449

RESULT 13
AE005123
LOCUS AE005123
DEFINITION Halobacterium sp. NRC-1 section 154 of 170 of the complete genome.
ACCESSION AE005123
VERSION AE005123.1
KEYWORDS GI:10581826
SOURCE
ORGANISM Halobacterium sp. NRC-1.
Halobacterium sp. NRC-1.
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halobacterium.
REFERENCE
AUTHORS Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M., Shukla H.D., Laaky S.R., Balliga N., Thorsson V., Shroana J., Swartzell S., Weir D., Hall J.D., Dahl T.A., Welti R., Goo Y.A., Leitner S.B., Keller K., Cruz R., Danon M.U., Hough D.W., Maddocks D.G., Jablonski P.E., Krebs M.P., Angelvine C.M., Dale H., Isenberger T.A., Peck R.F., Pohlschod M., Spidich J.L., Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L. and DasSarma S.
TITLE From the cover: genome sequence of halobacterium species NRC-1
JOURNAL Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)
PUBMED 11016950
REFERENCES 2 (bases 1 to 11202)

TITLE Direct Submission
JOURNAL Submitted (14-JUL-2000) Institute for Systems Biology, 4225
AUTHORS Roosevelt Way NE, Seattle, WA 98105, USA

FEATURES		Location/Qualifiers
source		1..11202 /organism="Halobacterium sp. NRC-1" /strain="NRC-1" /db_xref="taxon:64091"
gene		76..1281 /gene="meta" /note="VNG2420c"
CDS		76..1281 /gene="meta" /note="Meta" /codon_start=1 /transl_table=11 /product="homoserine O-acetyltransferase" /protein_id="AAG20507.1" /db_xref="GI:10581827"
gene		/translation="MAALAIVAHNTVAKNYOYCCTDGGPMPTTELDGFFRCGDTV DSLTVYATRCGDTGSNAVLVGHALTGSHVHGRADGDRGMPPVIGRGRPIDTTEY FYVCANVPSCHGTGPPADGPGDPCGTAFPPVTPMDTRCQRRLDTLGVRLHAV VGGVSGMNAIDMARRYDDVHRIAIVATAPRLAQCLGLNARRAIRKGNMNDGD YGGAHPRGGLALRQLGHVYLSKASMADEKFRRTATTDALPADPAAEFFATDVE SYLDQADTFVARFPANSYLLRLAMDADDLAAGRPVADALAFDGDALVLSFTGDM HFTAQASATVADAFRAGGAAGVAHHVVDSDYGHDAFLVDSAAVPRPLADFLAAGIDGA ATDATATGCV"
gene		1351..2646 /gene="hal" /note="VNG2421c"
CDS		1351..2646 /gene="hal" /note="hal" /codon_start=1 /transl_table=11 /product="O-acetyl homoserine" /protein_id="AAG20508.1" /db_xref="GI:10581828"
gene		/translation="MGHDGLHTNSVHAGORVATPGARAPLYOTTSVAFPSDAA GCFALERGCYITSRLMNPFVETLQRLALBEGCVAVATASGMAALDLATFLIARAGD SVAAASDLTGTVITLTHSAQRGVDTTFVDYADADADADATVAVLVEVTPGS LTPDLERLATDANGVPLVDNFTFATPALATPIDHGADIVHSTTKIHGAGTVG GALVADGSPDMADHADYPEIQAENPAGVFTDFRFGDAFTVAIAAGLDLINOQ SPEDAMOTLOKLETLPLRMOCHRNAQLVAEHLRDPNVSMVNYGALADHDHNAAT YLDGCGMLTFGVDGYEAAQSVTEETLISLANVGDAKTLVHPASTHQQLPTE AORAGVPRPMRVSGIEDPADIADLETALAEAVGSA"
gene		2715..5753 /gene="gldC" /note="VNG2422c"
CDS		2715..5753 /gene="gldC" /note="gldC" /codon_start=1 /transl_table=11 /product="glycolate oxidase subunit" /protein_id="AAG20509.1" /db_xref="GI:10581829"
gene		/translation="MATRDLRSDSDPDDADYDTSVERPGLVDDLRERVDGR FDSYRQLYATDASTYELPVGVPPTSTDVDAVAVSTCADRGIPVLPFGGGTSLAQ AVNEAVVLDLFRHAMNGFVGDEAEFAEQPCITLAALNDELALHAKAPDPAMGDK SVLGAIGNNSTGASLQYEKADGYLESAEVLADGTVTTLGMDVTDIDALADPDG LEARIYADLRLLDEADRIIRRFPPDLKRNNGYNDRLDLAEAGAVANARLLAGS EETIGITAEVSLIEPVAETAVAMLTVDVADQADVAPVLDDHPSAEVMDVDFLD LARDFDERADYVGLPDDGTSTLVEFYADSPAERARRADILDRBPATPRAADPPG AATTPDAVHADLEAYDDDRQAHFWMKRAQLPILSTSDKEHYPVIEDVAIPADH LPAVADFOAILDDLDTVASYYAHGPGVLPVLTADGVAKMAEIAIRVLDV VYGGAVSGEHDGRASQMNKRFEYGEDLMASFRRTKAFDPDWILNPVCGDAAE IDMTDLHRAVPDEDFDAGFPALNMDTENGFRGMTLCCAGCGTGHOOHTGSCMPT VYAAENETLSTGRBANMLBSAMGNLPPADLDEPLTVEYLDLCVCGKCKGLCCPSGD MAKLTVEYKHASHQADGVGVEDRLRGHTALFHLGSALAPVSNMATALGAGALAEKT LGIAERDLPTFRRESLRAMFDRAGSGRVPEAAEHTVLLADPPTTNTSTVAVGAAY RVLEAGVIVALPDGVTSGRPAYSNGLIDRTKRLAREVVAALAPRVADGMVYVYER SDAVYODDYHDLISAPAAVAAVNTYGVCEYLDLRHGDAAAGFDAGDGESELYHGICH OKAQRDHHAGVLRAGYEVDLPDSTCCGMAGSGYEAEHY SMSQAIADTLADQVAD SDADVTATAGSSCSRSLSDRDDVDTDGAPRPHVOKLADAL"
gene		complement(5763..6470)
gene		/gene="serB" /note="VNG2423c" complement(5763..6470)
CDS		/gene="serB" /note="serB" /codon_start=1 /transl_table=11 /product="phosphoserine phosphatase" /protein_id="AAG20510.1" /db_xref="GI:10581830"
gene		/translation="MLCWSAPKVCYCMKCVYAGRRATPGMTLVAFEDGTLAESMLDR IARAGGDEVAATITERMARGELSTADSLREPAQVLVAGIPESAAAAYVDGRLRAGAC DLVAKLRDGGVAVVLTGTFKPGVAAAPDAGVADGVGNLVADELGAEGPL VEGTRKDALRDCAEAGTTPAAAVAVVGADANDVPMLDAAAGTAIGVDPKPGVADCHDT VSMDALGRVLDDHGA"
gene		6555..8138 /gene="serA1" /note="VNG2424c"
CDS		6555..8138 /gene="serA1" /note="serA1" /codon_start=1 /transl_table=11 /product="phosphoglycerate dehydrogenase" /protein_id="AAG20511.1" /db_xref="GI:10581831"
gene		/translation="MRVLVTDPIADAGLRLRDAGHEVTTAYDATGALLDVAUSDAA LIVRSQTAVTDVAFAPDLVIVARAGISVDNIDIDATDHDGVANAPAGVNRAAE HTVALAPFAARSIPQAHARLDGSMAKDDYLGTELSGRTIGVVGARGOEVATRLDS LGMDLVADPYPIGEDRAOIGAEVDIETCVARADPLTIHVPLDGETGLIGEARLLR MDGCVIVARCGVYDEDLADRAODGVYAGALADLVFRTEPLPAASPLHADSITTP HIGASTKAQENAVATDADQVVAALAGDPVVALNAPSASERSAFDRIRFVDLSTAG TVAALFPRIERIVDVTYQGEVADDEVALVYAAQOGAFAGLEQVNVNAPRAVEER GIAVETENHRESEAVOSLVSIVGNGDAELTVSGTLFAGDDPRLVEIDGFVEAPNG HMLVARNHDTPGVIGTIGVLTGCTAEVNIAGMNFAREANGGALTYYNLDADVPARALD ELAGDRLVDVYSIELNCE"
gene		8200..8643 /gene="act" /note="VNG2426c"
CDS		8200..8643 /gene="act" /note="act" /codon_start=1 /transl_table=11 /product="acyl-CoA thioester hydrolase" /protein_id="AAG20512.1" /db_xref="GI:10581832"
gene		/translation="MGSTSDTFMENRERYOPDPTNNVYASAHGCVYKMMDEIGAMAAM RKAGTCVTARINSIDFERPVPGDICIYQATAYTGHSTIVRLRARREDRSGETE PTTDSIFVFVAVDMDMOPVAVPDLVGSERCKELREALAGHDE"
gene		8825..9886 /gene="dld" /note="VNG2429c"
CDS		8825..9886 /gene="dld" /note="dld" /codon_start=1 /transl_table=11 /product="D-lactate dehydrogenase" /protein_id="AAG20513.1" /db_xref="GI:10581833"
gene		/translation="MISALTADIHITNTKAGFAFTTLMGLVALLOYPSGRSLDLSRK TLVVLGASGSGFVVALNARSYAALLVGAALVGTGAGLYPTAARALVADLYSTQGR AGFLHTASGDLGMSAAGIAVALALLSWRYAFVPICTLAVALIALVWSRESYVE RSLSLVRDPAALLSGTOERMLVAVTYLAFWQWIAFLPFTDGEFEFVAVGTTVA RGALFAVGAIAKPTGTLSDPTSRALAVGALVGAALSAVYVPSVPAANGAVYVF FAGLILAFVPMOSTLMDAFPPESAGDUGAMTIVYIGLALGPTTYGVAVSTANETIA EWGLGALLVAIVIIARVT"
gene		complement(9890..11137)
CDS		/gene="thrc1" /note="VNG2430c" complement(9890..11137) /gene="thrc1"

```
/note="Thrc1"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AKG20514.1"
/db_xref="GI:10581834"
/translation="MTDLSLAGEETPACATDGVLMACVECGTILAPFADVVYRCPECSG
LLEARYADPGFEAFSGRGMVRYADALPVDAVSVQEGDTPLEYEAPDVAAAGVADVR
VHEGMNPTGSKDRGMVGVAVRLEGLDRLACASTGNTSALAAVGRATPVLYL
LPAGVAAKIQALQALHGARICEVDGNDPDCDVALELAAREAYLINSINPRLGEO
KTIAREILEOSTRTATGEMPRDRLVLYVNAAGNTALYKARELYAAGMTDDMPALIG
VOAGAAPVVEAVEAERGTERMTWVETRAVIRIGNPVNAKRALPGVETGTAVALV
SDDRITDAORMLAGDGIQVEEPASASVAGLLALRREGISADERVVCILTTGHLKLDPA
AAAAAGCTTTPPADADGVLDAI"
BASE COUNT      1670 a      4414 c      3626 g      1492 t
ORIGIN

Query Match      9.4%; Score 69; DB 1; Length 11202;
Best Local Similarity 43.8%; Pred. No. 0.3;
Matches 300; Conservative 0; Mismatches 385; Indels 0; Gaps 0;

QY 18 cgaatgcgcaagaagaaagaaacgagacgtgccccgagccgacgctcgagctgtgcgg 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7058 CGACCCCTTCATCGGCGAGACCGGGCGGCACAGCTCGCGCGAGCTCGATCGACATCGA 7117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 78 acgctggagacgcgcgagatgtgacgagggcgctacacgggtgatcaacaacgtatgggg 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7118 GACCTCGCTCGCGCGCGCGGATTTCTCAACGATCCACGTCCCACTCACCGACAGACCGA 7177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 138 cgcgcgaagaccgcgcagatgtgcatgtgaggttcggaatggaaacgggcaacttcacatcaacg 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7178 CGGCTTCATCGGCGAGACCGCAACTCCCGGATGAGACGGCGGCTACGTCAACGTCCG 7237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 198 ggcgcgacacgaacaagacgtgacgcgctacacgctactacttgcgtggcgca 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7238 CCGCGCGCGCGCTGCTGACAGAGACCGCTGCGCGACCGCGCCACAGACGGCTCATCGC 7297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 258 ctggggcgccctgcagacgaatcgtggatgtccgcgcgcgcgtgcagagctgtctcgacgt 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7298 GGGGGCGCGCTTCGACGTGTTCGACGGAGCGCGCTCCGCGCGCGCTCCCGCGCTCCACGA 7357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 ggcgaagaagctggaacgtcaacgcgacatcgaagcgcgctgtgaatgcgcgcctcaagacat 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7358 CGCGACACACATCATCACACACGACACCTGGGGCGCTCCACAGACCGCGCCACAGAGAA 7417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 ctggttcagtcctcgtcaagaattccgcaacgctacagcgcgcgccgagctgagtgat 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7418 CCGCGCACACGACACCGCGCACAGGTGCTCGCGCGCTCCCGCGCGACCGGTGCTGAA 7477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 438 ctggttcgaactggaacgcgcgtgtgctgcggcgcgcgcaacgcgcgtgcagccgttggaact 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7478 CGCGCTGAACCGCGCCCGACGCGCGCTCGCGCTTCGACCGCATCCGCCCTTCGTGCA 7537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 498 ggcgcggcgcaacttgggaagctctgtatgtgcgaacttggagacttgatataatgcctaacg 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7538 CCGCGCGAGACCGCGCGACCGCTCCCGCGCGCTTCGACCGCCCGCATCGAGCGCGT 7597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 558 ggcgaagaagcccaacacgctcgttgagcgagctggaagcctcctcaacgaacg 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7598 CGACGTCACAGTACCAAGGAGAGGTCCGCGACGAGACGTCCGCACTGTCACAGCGCGCGC 7657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 618 ggtgcgcgcgcgtacatccgcgcgagtgatctgtcatgtcggtggagagacgggtcga 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7658 CCAACAGCGCGCTTCGCGCGCTTCGAGTGGAGTCAACGCGGTCAACGCCCGCGGGT 7717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 678 actcttggagggcggtggcggtctgtc 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7718 GGCGGAGGAGCGCGCATCGCCGCTG 7742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
LOCUS      AF335723      4262 bp      DNA      linear      BCT 11-FEB-2001
DEFINITION Burkholderia pseudomallei formate hydrogenlyase subunit 4 and
ACCESSION  AF335723
VERSION    AF335723.1  GI:12744948
KEYWORDS   Burkholderia pseudomallei.
SOURCE     Burkholderia pseudomallei.
ORGANISM   Burkholderia pseudomallei.
            Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
            Burkholderia, pseudomallei group.
REFERENCE  1 (bases 1 to 4262)
            Steiner,B., Meyer,R., Bowen,M. and Morrill,W.
            Random sequencing of Burkholderia pseudomallei strain G9313 for
            clinical PCR development
            Unpublished
JOURNAL    2 (bases 1 to 4262)
            Steiner,B., Meyer,R., Bowen,M. and Morrill,W.
            Direct Submission
            Submitted (12-JAN-2001) RRAT, Centers for Disease Control and
            Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
            Location/Qualifiers
FEATURES   1..4262
            source
            /organism="Burkholderia pseudomallei"
            /strain="G9313"
            /db_xref="taxon:28450"
            complement(89..1039)
            /codon_start=1
            /transl_table=11
            /product="formate hydrogenlyase subunit 4"
            /protein_id="AKK06856.1"
            /db_xref="GI:12744950"
            /translation="MVASAGILSQTEILVALAALPLTGVNOCRAMLQNNRPAISW
            QPYRMILKLFNKESVAVAGAREPLFAGVAVVWAAALACAIPTLSTELSPSPADAI
            ALVGFALARVALSLAAMDIGTAFGLTARREMLGLAEPLALVSAISITSTL
            LHSIVAAIGRELAITYPSLAEGIAFTVLSAENRCPVANDPATILEMTHEALILE
            VSGRIALMEMASIKIFAYSCIGIALMPMCISANRGLPDLALLALPLFLKLVNGA
            ALAVETTNARKRLFRVPEFLASATFLVIGLVHFLGA"
            complement(1033..3039)
            /codon_start=1
            /transl_table=11
            /product="hydrogenase-4 component B"
            /protein_id="AKK06855.1"
            /db_xref="GI:12744949"
            /translation="MEGLLKCULFVALAMIALCAGLALNRRGTVAHGLFPLGAL
            AGLVLAALAGLAVFAGPSAVLPLGLPLDPLRPLRPLDPLSAFLVGLVTVAGVAFSS
            GYFRKGEAGPLICFEYHVCVAVSLVLDVADVCAPVWATLTLTSAFTVMTNRI
            AEIRRAAFVFLISHVGLALILFCGLAQAGGDTYFANMRIOHLDGLAASAEVIAL
            VGEFGAKGIFPLHVMPLPAHPAPSPVALLSGFVLKGVIGILRTVDLHLVQVAM
            GYLMALGLFGLALGAVSTIOTDMKRLIAYSSIDNGLMVCVSLGTLTLFLGYRLPSL
            MALSTALYLQIVAHACFSLFLCTGAVLHATGRNLRUGGLIRMPWTAMALV
            VAAAGLPPSSGSEVSEMLLVOSFLTGPGLPDSLVGLVPLVAAVLAALPAALAGAMVK
            FEGIVFLGRPREAKLERADASPWERIAFGVMAASVLLGLAPVOYVVLVLAALTAALV
            GAGIGLTHNGMLLPAPVSTARASVYRPALEMLFEVACALAMILLVRFVHGRLRAPV
            SCGNFVTRARMODTAEFGOPRIREITFEFFERLIDPSFDROPVRYMSVADRTWHLL
            YEPVIALNRAVALYGRLOTGRIALYITFSVLIALLTIVRRW"
BASE COUNT      744 a      1502 c      1477 g      538 t
ORIGIN

Query Match      9.4%; Score 68.8; DB 1; Length 4262;
Best Local Similarity 47.3%; Pred. No. 0.41;
Matches 208; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 41 cgaagcttagcccgagaccgctgagcgtggaacgcttggaacgcgcgagatgtgg 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1404 CGCAGCAGCGGACGAGAGACATGACAGCGCGCGCATGTAGCTCGCGCGCGGCTGG 1463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 ccggggggcgctacccgggtgatcaacaacgtatggggcgcggaagaccgcagtgcatgtg 160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1464 ACACGGGCGCGAACACACACCGCTGATTCGTCGCCGAGTACCGCGCGACAGCGG 1523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 161 agtctgacttgaaacgggcaacttcagatcaacagggcgatcaacgcaacggcaaca 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14
AF335723

Db 1524 CGCGCGTGAACGATGACGACCTTACGACTGACGAGCGGCGGAGCGCCGACGACGACG 1583
OY 221 acgtggccgcctatccggccatctacttcgggtgctcagctggcgccctgacagcaatt 280
Db 1584 TCCGGGCGCGGACACGACGACGATCGGCTCCGCGGCTCGCGGCGCGCT 1643
OY 281 cgggattgcggcgcgcgctgcagagctgtcgcagctgcgcagagctgcagctcagc 340
Db 1644 CGAGCTTGCGCTCGCGGCGGCGGCGGAGAACGAGAGCGGAACTGACCATCGCGT 1703
OY 341 ccgtcagcagcgccgctggaatgcgcgcctcagacatctgttcagctcccgctcagcaat 400
Db 1704 AGCCGGCGACGCCCGCTGACGCGCAGCGCGCGGACGAGCGGACCGCTCATCCCGA 1763
OY 401 ccggcaacggctacagcgcgcgcgagctgatactgtcgtgaactggaacggcgcg 460
Db 1764 GCACCGAATCCGGCGCGCGCGGTGAAACAGAAAGCTCTGACAGAGACGCAATTCGACA 1823
OY 461 tgatgcggcgcgcgagcgcg 480
Db 1824 CGAAGCGCTCGAGCGCGC 1843

RESULT 15
AE006992 14483 bp DNA linear BCT 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 78 of 280 of the
DEFINITION complete genome.
ACCESSION AE006992 AE000516
VERSION AE006992.1 GI:13880691
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacterium
Mycobacterium tuberculosis complex.
1 (bases 1 to 14483)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,M.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Kouri,H.,
Gill,J., Mklula,A. and Bishai,W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 14483)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,M.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Kouri,H.,
Gill,J., Mklula,A. and Bishai,W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
FEATURES
source

gene
CDS
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
complement(66..560)
/gene="MT1111"
complement(66..560)
/gene="MT1111"
/note="similar to SP:P21346; identified by sequence
similarity; putative"
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcription elongation factor GreA"
/protein_id="AAK45367.1"
/db_xref="GI:13880692"
/translation="MPTQVTVLTQSHRRLKAEELDQLIANRPVIAEINDREEDL
RENGTHAAREEGQOEARIQLODILLSNAKVGAPKQSGVALPGSVVYVYNGKSD

SETFLIATROEVSQGLEVYSPNSPLGALLIDAKVGETRSTYTPNGSTVSTLVSAE
PYHS"
complement(746..1207)
/gene="MT1112"
complement(746..1207)
/gene="MT1112"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45368.1"
/db_xref="GI:13880693"
/translation="MHPAGVNPQMTHTPIPPDARYGRPLSRARRRVAALGVVA
AAGIVAVIGYORISYSAVTSGLVGRVLDDELTASVTISVRSDESPVACIVRAT
NSENQGRRLVPPSEATVQVTTVKSQPPVMAVDYCGTEVPSYILP"
1282..2148
/gene="MT1113"
1282..2148
/gene="MT1113"
/note="identified by PID:699274; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ImbE protein"
/protein_id="AAK45369.1"
/db_xref="GI:13880694"
/translation="MSELRMAVHAHPDESSKGAATLARADEGHVLYVLTGGER
GELINPAMDLPVHGRHAEIRDEMTKAEITGEVHTWGLFVDSGLPKDLPPLPD
CFARVLPVSTELVYVREFPHVVTYTDENGGVPHPHIGHOVVAAYAADPDC
RPDAGPPTVSKLYVHGRHMOLODERARHQGRFPQMLAYNDPDLTSLR
VTRVDECSKTFESQDDALAHATQIDPNAEFFRAPLAWOERLPTFEELASRLPAR
PPELEFAGIEP"
2145..2411
/gene="MT1114"
2145..2411
/gene="MT1114"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45370.1"
/db_xref="GI:13880695"
/translation="MNQILSYIAEGCGPNTGPRDKASPVGLVLYLVLTPLVR
SMQQLKVPKSPDRHPELDPADEGTDRDGPAPPPHESG"
2398..4419
/gene="MT1115"
2398..4419
/gene="MT1115"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK45371.1"
/db_xref="GI:13880696"
/translation="MSPANPSCGNTLALATSPYLROHADNPVHWOQWTPQALAEAAAR
AVPILSVGAACHMCVMAHBSFDDDEVAAMNAGFVCIKVDEREPPIDVYNNAT
VALTGGGWPMTCTFLPNRPFPCGTYVYKKAFLDLSAISTWERRAEVQASDHI
AAELRSMASGIPGCGPEVAPELDCDVAVALREODPAHGFAGKPFEPSSALELMR
HYERTSPALAEVAVARTGNAMARCGIYDLDGGFARVSDGAVVVPHEPKMYDALI
LNAVHARRTGDPRLARRVAQOTARFLDELISKAPADMTSSLADADRGCSITYV
TPVQLTEVIGDGDGRWAAEVFGTEGTEHGTSTVQLCPADPDAAARDVRALLVA
RLARAQPARDKRVYTSWNGLATITALEASVALDDPALAARCATRLDLHVVDGR
RRASLGCVGDSAAILEDHAMLATGLALYQLTSEGAVLTATGTLDTVAHFVGPOR
PGMWPTADDAERLMRPSDPLDGATPSSASSIAEALTFAGHVGARARERYWOLAD
TLRAHVALARAPRSAGHVALAENAVRSGASSTIAEALTFAGHVGARARERYWOLAD
VGAAGSSALLVGRDVRVAGADAAYVCRGVCDLPTVSAELATALGVCP"
4429..4548
/gene="MT1116"
4429..4548
/gene="MT1116"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

gene
CDS

/product="hypothetical protein"
/protein_id="AAK45372.1"
/db_xref="GI:1380697"
/translation="MAPVREIRDAVSACRVAIFTLATALGVPQONTNTNHM"
complement(4518..5246)
/gene="MT1117"
complement(4518..5246)
/gene="MT1117"
/note="similar to PID:887849 GB:U00056 P.D:1789266;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Hly-III family protein"
/protein_id="AAK45373.1"
/db_xref="GI:1380698"
/translation="MSGQADPRATTAEARTPAHAHLVEGVARYLTKRRRGTHYVS
ACTAYLAGSLVAVSMAHGSAKAGITLATYATTAITTFYSATIHRYNMKSATARKMM
KADSHSMFEVFIAGSYTFPALPAPHDGRVLISWGGAIAGILMKCPAARSYG
VLDLLLGVAWYATATILHNAGVALVLEFGALYSIGSILYAVBWPDPWPTTEGY
HEEFNACTVVAICHIYIAMWEFV"
5357..6145
/gene="MT1118"
5357..6145
/gene="MT1118"
/note="similar to SP:P20182 PID:1196916; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="undecaprenyl diphosphate synthase"
/protein_id="AAK45374.1"
/db_xref="GI:1380699"
/translation="WEIIPRKKEPYLRLYELRLROGLASKSDLPHRIIVLCQNR
NARSGVDVSYGYRMGAAKLAEMLRMCHEGIELATYYLLSTENLRDPELALIE
ITIDVVEECIKANHMRSVRATGDGLIGEPPARLRGVESTPSAFHYNVAVXGSG
REIIDVARALKSELANGATAEELVYATVEGISENLYTSGCDPDPLVTISEORT
SCFLMOASYSSEMFTFAHWPAFRHVDFLRALRDY SARHRYSR"
6283..8591
/gene="MT1118.1"
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
8768..9088
/gene="MT1118.2"
8768..9088
/gene="MT1118.2"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45375.1"
/db_xref="GI:1380700"
/translation="MPCVGSDRRFPNUAVAVEAICENTANTSGOPDDIVRTSGEDR
LSGHNGPGCVGSRRRLRALNDYSTPHASIFYPPPIKSDGITHASRLAVESVDALAG
RVEL"
9223..9675
/gene="MT1119"
9223..9675
/gene="MT1119"
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="PE family protein"
/protein_id="AAK45376.1"
/db_xref="GI:1380701"
/translation="WKGGLMSYMIATPALTAAATDIDIGSAVSVANAAAATTC
VLAAGDVLAAIALRLFNALEEYHALSAOYAATQTLFVRTLTGGCGVFFRRRKOCV
TAAERRAAGAGRRORRRRSSDGOWRLRQQRHFCGGGCPFEROHSEHR"
9692..9859
/gene="MT1120"
9692..9859
/gene="MT1120"
/note="similar to GB:AL123456; identified by sequence
similarity; putative"

	Query Match	8.6%; Score 63.2; DB 1;	Length 14483;
	Best Local Similarity	46.8%; Pred. No. 1.8;	
	Matches 304;	Conservative 0;	Mismatches 338; Indels 8; Gaps 3;
QY	81 cctggagcgcgcgcatgtgycgcggggcgccgtaccgcggttgatcaacaacgatgtggc-g 139	/codon_start=1 /transl_table=11 /product="PE family protein" /protein_id="AAK45377.1" /db_xref="GI:13880702"	/translation="MTTASATASSGVDSGIAATVAVASQWDGYVANVTITGTGRSF
DB	10132 CTAGTTCTGGGACGACTACGTCACCGGAGGCTACGTGTGCAGAACCAACGCTGGACAATA 10191		
QY	140 cggagacgcccccagtcgcatctgaagtgcgaactygaaaacgycgaacttccagatcacagg 199		
DB	10192 CCCCCGCCGGGAGACCGTCACGTCACGCAACCCGGGTTTCACCTTCACCAGAGATGAACG 10251		
QY	200 ccgatacagacaaacygcaaacagtgtgcgcgcctalcscgacatctacttcggtgcacct 259		
DB	10252 GTGTGTGCCCAACCAACGGCGCCGCGTGAGTTACCCCTCGATTGTGCGAGGGCGTGCACCT 10311		
QY	260 gggg-gcctcgaagcagaactctcggaattgcgcggcgcgctgcagagctgtctcgaactg 318		
DB	10312 GGGGGCACCTCCTCGGTGGGACCAACCTCGCTACTGAGTGTGGCCAAATTTTGTGGCGC 10371		
QY	319 cgacgaagcttgaagctcgcgcgataccgaacggcgcgcttgaaatgycgcgcttagacatc 378		
DB	10372 CCGACCAGCATGACTACACTAACCAGACGCGGGGATGTGGAGCGCTCTCTACGACATC 10431		
QY	379 tggtaagtcgccgttacgaattccgagcaacggctacagcpgcgcgccgagctgtatga 438		
DB	10432 TGCCTGGAATTCACAACCCAAGACGACGGGGTCAACAGCA-----GAGATCATGTATC 10485		
QY	439 ttgcttgaacttgtaaaacggcgcgctgaatccgcggcgcgagcgcgtgtgccaacgltgaactg 498		
DB	10486 TTGTTTCACACACAGGGGCTCACTTAAGCGCGGTGGCTCCCCTGGTGGGAAACCAACCAATC 10545		
QY	499 gcgcggcgccaaacttgggaagtcttgatctccggaacttgggaacttgaatatcatgcctlacgg 558		
DB	10546 GAGGGCAGAAGACTTGTGTGTGTGGATGGACCAACAGCGCATGAAACAGCCGATGGCTTAT 10605		
QY	559 cgacgaagcccccaacagctctgtatgaagcagctgcgaactgaagagcttcatcgaagacgg 618		
DB	10606 GTGCGGACCGAGCCGATGAGAGGTTCGAGAGCTTCGACAGTGATAGTTTTGTGTGACCAACAC 10665		
QY	619 gtgcgcgcggtacatccgcgcgcgagttgtatcttgcatlcggttgcgttgaagacgggtcga 678		
DB	10666 GCCACCATGAGACCGCATCCAGACTCGTGTGTACTCACAGACATCCGGGCGGCTTGGAG 10725		
QY	679 cctctggagagcgcggtccggtctcgaagcgccgaatttccgttaagcgt 728		
DB	10726 CCTCTGAGAGGAGGTGTGGTGTGGGGGTCATTCTCTGCGCAAAAGT 10775		

Search completed: August 19, 2002, 12:45:06
Job time: 10450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:53:47 ; Search time 544 Seconds

(without alignments)
2310.260 Million cell updates/sec

Title: US-10-003-759-3_COPY_52_783

Perfect score: 732
Sequence: 1 tgcgactggtccttcctccga.....atttccgtaacggtgcag 732

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq.032802:*

1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.8	18.0	1116	20	AAV72359
2	131.8	18.0	1116	21	AAZ57029
3	131.6	18.0	1715	21	AAZ57031
4	57.6	7.9	1983	22	AAA54513
5	53.6	7.3	77536	21	AAA14651
6	49.8	6.8	756	17	AAT16766
7	49.4	6.7	1404	21	AAA38763
8	49	6.7	756	14	AAO43032
9	48.8	6.7	985	19	AAV64548

10	48.8	6.7	985	19	AAV44439	Mycobacterium tube
11	48.8	6.7	985	20	AAZ19349	M. tuberculosis an
12	48.8	6.7	985	20	AAZ19137	M. tuberculosis re
13	48.4	6.6	12152	22	AAZ08699	M. carbonacea DNA
14	48	6.6	4403765	22	AA199683	Mycobacterium tube
15	47.8	6.5	24379	18	AAT93095	Streptomyces freno
16	47.8	6.5	24379	18	AAT25925	Streptomyces roseo
17	47.6	6.5	30001	19	AAT61016	Total DNA sequence
18	47.6	6.5	30001	20	AAZ05110	S. aureofaciens DN
19	47.2	6.4	1908	16	AAO99364	S. lividans protea
20	47.2	6.4	1908	20	AAV84065	Tripeptidyl aminop
21	47.2	6.4	1908	21	AAZ61403	cDNA sequence enco
22	46.8	6.4	15872	21	AAZ61283	S. venezuelae vep
23	46.8	6.4	20394	22	AAZ24892	Plasmid biosynth
24	46	6.3	44377	18	AAT78508	Platenolide synth
25	46	6.3	44377	18	AAT80414	Platenolide synth
26	45.6	6.2	744	18	AAT74073	Mycelophthora the
27	45.2	6.2	1291	20	AAZ87940	Mycobacterium tube
28	45.2	6.2	2852	19	AAV64558	M. tuberculosis im
29	45.2	6.2	2852	19	AAV44449	Mycobacterium tube
30	45.2	6.2	2852	20	AAZ19359	M. tuberculosis an
31	45.2	6.2	2852	20	AAZ19147	Nucleotide sequenc
32	45.2	6.2	77536	21	AAA14651	Non-reducing sacch
33	44.8	6.1	2268	21	AAA10501	Rice 1-deoxy-D-xy
34	44.8	6.1	2618	21	AAA38760	Non-reducing sacch
35	44.8	6.1	3252	21	AAA10505	Mouse Sox1 cDNA.
36	44.6	6.1	2376	20	AAZ16153	Polyanion brachys
37	44.6	6.1	2712	19	AAT94214	L05390 cDNA clone.
38	44.6	6.1	5224	22	AAZ90079	S. venezuelae macr
39	44.6	6.1	11220	21	AAZ87298	Sequence encoding
40	44.6	6.1	12588	15	AAZ063293	S. venezuelae pik
41	44.6	6.1	36778	21	AAZ87318	Nucleotide sequenc
42	44.6	6.1	37948	21	AAZ87285	Recombinant cosmid
43	44.6	6.1	38506	21	AAZ75633	H. virescens inorg
44	44.6	6.1	38506	21	AAZ56001	
45	44.4	6.1	867	22	AAH48828	

ALIGNMENTS

RESULT 1	
AAV72359	standard; DNA; 1116 BP.
XX	
XX	AAV72359;
XX	
DT	28-JUL-1999 (first entry)
XX	
DE	Actinomyces sp. 36KD cellulase DNA.
XX	
XX	Cellulase; detergent; animal feed; nutritional value; textile;
KW	stone washing; texture modification; appearance; cellulosic fabric;
KW	pulp; draining; paper; baking additive; starch treatment; grain;
KW	high-fructose corn syrup production; ethanol production; fibre reduction;
KW	milling; ss.
XX	
OS	Actinomyces sp.
XX	
PN	WO9925846-A2.
PD	
PD	27-MAY-1999.
XX	
PF	18-NOV-1998; 98WO-US24649.
XX	
PR	24-JUN-1998; 98US-0104308.
PR	19-NOV-1997; 97US-0974041.
PR	19-NOV-1997; 97US-0974042.
XX	
PA	(GENW) GENENCOR INT INC.
PI	
XX	Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;

[illegible]

RESULT	3
AA257031	
ID	AA257031 standard; DNA; 1715 BP.
XX	
AC	AA257031;
XX	
DT	19-MAY-2000 (first entry)
XX	
DE	DNA sequence of an expression cassette containing cellulase 11AG8.
XX	
KW	Cellulase; Actinomycetes; detergent; feed additive; textile treatment; pulp; paper; cellulase 11AG8; ds.
XX	
OS	Streptomyces lividans.
OS	Synthetic.
XX	
FH	Key
FT	promoter
FT	Location/Qualifiers
FT	1..406
FT	/*tag= a
FT	/note= "GI promoter sequence"
FT	407..548
FT	/*tag= b
FT	/note= "Cela signal sequence"
FT	549..1564
FT	/*tag= c
FT	/note= "mature 11AG8 sequence"
FT	1576..1715
FT	/*tag= d
FT	/note= "11AG3 downstream sequence containing terminator"
XX	
XX	
PN	WO200009707-A1.
PN	
PD	24-FEB-2000.
XX	
PF	28-MAY-1999;
XX	
PR	99WO-US11971.
XX	
XX	98US-0104308.

[illegible]

```

QY      712 gatttttcgtaacggtgca 731
          |||||
Db      1187 tcgtctctcctcgcggtgaa 1206

```

RESULT	4
AAA54513	
ID	AAA54513 standard; DNA; 1983 BP.

AC AAA54513;

DT 11-APR-2001 (first entry)

DE Fructan exohydrolase (FEH) coding sequence.

KW Fructan exohydrolase; FEH; transgenic plant; recombination.
 KW transgene; gene expression; detergent; detergent additive;
 KW oral care composition; ds.

OS Zea mays.

PN WO200068402-A1.

PD 16-NOV-2000.

PF 08-MAY-2000; 2000WO-EP04226.

PR 06-MAY-1999; 99BE-0000329.

PA (LEUV-) LEUVEN RES & DEV.

PI Van Den Ende W, Van Laere A, De Roover J, Michiels A;

DR WPI; 2001-007401/01.

PT Novel DNA molecules encoding enzymes having fructan exohydrolase activity for use in transgenic plant production, dental care compositions, and in detergents

PS Claim 9; Fig 4; 45pp; English.

Transgenic plants such as *Clethra integrifolia*, *Cynara scolymus*,
Helianthus tuberosus, *Scorzonera hispanica*, *Oryza sativa*, *Zea mays*,
Triticum aestivum, *Triticum durum*, *Hordeum vulgare*, *Secale cereale*,
Avena sativa, *Sorghum vulgare*, *Panicum pruriens*, *Lolium temulentum*,
Dactylis glomerata, *Pennisetum americanum*, *Allium cepa*, *Agave
 americana*, *Agave azul*, *tequilana*, *Sorghum bicolor* and *Panicum
 miliaceum* transformed with a vector encoding a fructan exohydrolase (FEH)
 enzyme are useful for the recombinant production of FEH or other
 polypeptides having FEH activity. The FEH polypeptides produced
 are useful in detergents or as a detergent additive and in oral
 care compositions.

SD Sequence 1983 BP; 357 A; 614 C; 683 G; 329 T; 0 other;

Query Match	7.98;	Score 57.6;	DB 22;	Length 1983;
-------------	-------	-------------	--------	--------------

Matches 213; Conservative 0; Mismatches 1

Matches 213; Conservative 0; Mismatches 199; Indels 6; Gaps 3;

[illegible][illegible]

RESULT	5
AAA14651/c	
ID	AAA14651 standard; DNA; 77536 BP.

AC AAA14651;

DT 08-AUG-2000 (first entry)

DE Nucleotide sequence of the FK-520 biosynthetic gene cluster.

KM F520D polypeptide synthase; PKR; gene cluster; immunosuppressant
KM Streptomyces hygroscopicus var. ascomycesus; immunophilin;
KM KF-506 binding protein; polypeptide compound; transplant rejection;
KM graft-versus-host disease; warts; alopecia; universal;
KM autoimmune chronic active hepatitis; inflammatory bowel disease;
KM multiple sclerosis; primary biliary cirrhosis; scleroderma;
KM neurite outgrowth; nerve regrowth; Parkinson's disease;
KM Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KM peripheral neuropathy; ss.

05 *Streptomyces hygroscopicus*.

FH	Key	Location/quantity
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	CDS	complement (412..1836)
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129		

FT	FT	CDS
		/note= "fkbw gene"
		complement (2020..3579)

FT	FT	CDS	/note= "fkbv gene"
			3969..4496

FT	FT	CDS
1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30
31	32	33
34	35	36
37	38	39
40	41	42
43	44	45
46	47	48
49	50	51
52	53	54
55	56	57
58	59	60
61	62	63
64	65	66
67	68	69
70	71	72
73	74	75
76	77	78
79	80	81
82	83	84
85	86	87
88	89	90
91	92	93
94	95	96
97	98	99
100	101	102
103	104	105
106	107	108
109	110	111
112	113	114
115	116	117
118	119	120
121	122	123
124	125	126
127	128	129
130	131	132
133	134	135
136	137	138
139	140	141
142	143	144
145	146	147
148	149	150
151	152	153
154	155	156
157	158	159
160	161	162
163	164	165
166	167	168
169	170	171
172	173	174
175	176	177
178	179	180
181	182	183
184	185	186
187	188	189
190	191	192
193	194	195
196	197	198
199	200	201
202	203	204
205	206	207
208	209	210
211	212	213
214	215	216
217	218	219
220	221	222
223	224	225
226	227	228
229	230	231
232	233	234
235	236	237
238	239	240
241	242	243
244	245	246
247	248	249
250	251	252
253	254	255
256	257	258
259	260	261
262	263	264
265	266	267
268	269	270
271	272	273
274	275	276
277	278	279
280	281	282
283	284	285
286	287	288
289	290	291
292	293	294
295	296	297
298	299	300
301	302	303
304	305	306
307	308	309
310	311	312
313	314	315
316	317	318
319	320	321
322	323	324
325	326	327
328	329	330
331	332	333
334	335	336
337	338	339
340	341	342
343	344	345
346	347	348
349	350	351
352	353	354
355	356	357
358	359	360
361	362	363
364	365	366
367		

Accession	Gene	Score
FT	/note="fkbri gene"	5601.6818
FT		
CDS		

	/cug-	-
FT	/note=	"fkbe gene"
FT	6808,	8052
CDS		

ET	/lag= 1
FT	/note= "fkbf gene"
ET	8156 882A
CDS	

```
FT      g      /*tag=
FT      /note= "fkbg gene"
FT      complement 0133
```

```

FT      /*tag= h
FT      /note= "fkbn gene"

```

```

FT      /+tag= 1
FT      /note= "fkbi gene"

```

```
FT      /*tag= j
FT      /note= "fkj gene"
```

```

/*tag= k
/note= "fkpk gene"

```

```

FT CDS complement (12113..13150)
FT /tag= l
FT /note= "fkbl gene"
FT CDS complement (13212..23988)
FT /tag= m
FT /note= "fkbc gene"
FT misc_feature complement (13452..13662)
FT /tag= n
FT /note= "ACP6"
FT misc_feature complement (13761..14394)
FT /tag= o
FT /note= "KR6"
FT misc_feature complement (14517..15294)
FT /tag= p
FT /note= "ER6"
FT misc_feature complement (15438..16587)
FT /tag= q
FT /note= "dehydratase domain (DH) 6"
FT misc_feature complement (16587..17820)
FT /tag= r
FT /note= "acyltransferase domain (AT) 6"
FT misc_feature complement (17820..19053)
FT /tag= s
FT /note= "KS6"
FT misc_feature complement (19116..19326)
FT /tag= t
FT /note= "ACPS"
FT misc_feature complement (19464..20097)
FT /tag= u
FT /note= "KR5"
FT misc_feature complement (20241..21420)
FT /tag= v
FT /note= "DH5"
FT misc_feature complement (21420..22653)
FT /tag= w
FT /note= "AT5"
FT misc_feature complement (22653..23892)
FT /tag= x
FT /note= "KS5"
FT CDS complement (23992..46573)
FT /tag= y
FT /note= "fkpb gene"
FT misc_feature complement (24163..24373)
FT /tag= z
FT /note= "ACP4"
FT misc_feature complement (24997..26146)
FT /tag= aa
FT /note= "DH4 (inactive)"
FT misc_feature complement (26146..27430)
FT /tag= ab
FT /note= "AT4"
FT misc_feature complement (27430..28684)
FT /tag= ac
FT /note= "KS4"
FT misc_feature complement (28750..28960)
FT /tag= ad
FT /note= "ACP3"
FT misc_feature complement (29092..29740)
FT /tag= ae
FT /note= "KR3"
FT misc_feature complement (29869..31018)
FT /tag= af
FT /note= "DH3 (inactive)"
FT misc_feature complement (31018..32185)
FT /tag= ag
FT /note= "AT3"
FT misc_feature complement (32185..33439)
FT /tag= ah
FT /note= "KS3"
FT misc_feature complement (33505..33715)
FT /tag= ai
FT /note= "ACP2"
FT misc_feature complement (33823..34480)

```

```

FT /tag= aj
FT /note= "KR2"
FT misc_feature complement (34606..35749)
FT /tag= ak
FT /note= "DH2 (inactive)"
FT misc_feature complement (35749..37144)
FT /tag= am
FT /note= "AT2"
FT misc_feature complement (37145..38296)
FT /tag= an
FT /note= "KS2"
FT misc_feature complement (38371..38581)
FT /tag= ao
FT /note= "ACP1"
FT misc_feature complement (38677..39307)
FT /tag= ap
FT /note= "KR1"
FT misc_feature complement (39442..40609)
FT /tag= aq
FT /note= "DH1"
FT misc_feature complement (40609..41842)
FT /tag= ar
FT /note= "AT1"
FT misc_feature complement (41842..43093)
FT /tag= as
FT /note= "KS of extender module 1 (KS1)"
FT misc_feature complement (43144..43660)
FT /tag= at
FT /note= "ACP of loading domain"
FT misc_feature complement (43777..44629)
FT /tag= au
FT /note= "ER of loading domain"
FT misc_feature complement (44974..46573)
FT /tag= av
FT /note= "CoA ligase of loading domain"
FT CDS 46754..47788
FT /tag= aw
FT /note= "fkpo gene"
FT CDS 47785..52272
FT /tag= ax
FT /note= "fkbp gene"
FT CDS 52275..71465
FT /tag= ay
FT /note= "fkpa gene"
FT misc_feature 52362..53576
FT /tag= az
FT /note= "KS7"
FT misc_feature 53577..54716
FT /tag= ba
FT /note= "DH7"
FT misc_feature 54717..55871
FT /tag= bb
FT /note= "ER7"
FT misc_feature 56943..57575
FT /tag= bd
FT /note= "KR7"
FT misc_feature 57710..57920
FT /tag= be
FT /note= "ACP7"
FT misc_feature 57990..59243
FT /tag= bf
FT /note= "KS8"
FT misc_feature 59244..60398
FT /tag= bg
FT /note= "AT8"
FT misc_feature 60399..61412
FT /tag= bh
FT /note= "DH8 (inactive)"

```



```
FT      /tag- a
FT      /product- "1-deoxy-D-xyloulose 5-phosphate synthase"
FT      /partial
PN      WO200032792-A2.
XX      08-JUN-2000.
XX      02-DEC-1999; 99WO-US28587.
XX      03-DEC-1998; 98US-0110779.
XX      (DUPO ) DU PONT DE NEMOURS & CO E. I.
PI      Cahoon RE, Tao Y, Williams ME, Coughlan SJ, Weng Z;
XX      MPI; 2000-412338/35.
XX      P-PSDB; AAY97425.
XX      Polynucleotide encoding 1-deoxy-D-xyloulose 5-phosphate synthase enzyme
XX      useful for producing transgenic plants and for producing antibodies
XX      specific to which is useful for screening cDNA expression libraries -
XX      Claim 2; Page 67; 73pp; English.
XX      The present sequence is a putative coding sequence for the wheat
XX      1-deoxy-D-xyloulose 5-phosphate synthase enzyme (DXPS). Its protein is
XX      involved in the isoprenoid biosynthesis pathway. The cDNA was identified
XX      by sequencing a number of clones and then comparing their protein
XX      sequences to known proteins: this showed the sequence's similarity to the
XX      Capsicum annuum DXPS sequence. The DXPS gene and protein can be used to
XX      create transgenic plants which express the gene at either different
XX      levels or at different stages of development compared to normal, and to
XX      identify herbicides.
SQ      Sequence 1404 BP; 277 A; 482 C; 439 G; 206 T; 0 other;
```

```
RESULT 8
AAQ43032
ID      AAQ43032 standard; DNA; 756 BP.
XX      AAQ43032;
AC      07-SEP-1993 (first entry)
XX      Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.
XX      Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
XX      fibre; film; coating; triad sequence; collagen; mammalian; moulding;
XX      hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
XX      Synthetic.
XX      WO9310154-A.
XX      27-MAY-1993.
XX      04-NOV-1992; 92WO-US09485.
XX      12-NOV-1991; 91US-0791960.
XX      (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX      Cappello J, Ferrari FA;
XX      MPI; 1993-182496/22.
XX      P-PSDB; AAR37738.
XX      High mol. wt. collagen-like protein polymers - capable of being
XX      produced in unicellular microorganisms
XX      Disclosure; Page 39; 82pp; English.
XX      The sequences given in AAQ43030-34 encode examples of recombinantly
XX      produced DCP collagen-like polymers (CLPs) which consist of repeated
XX      tripeptide sequences selected from a wide range of GXY sequences.
XX      where X and Y can be any amino acid. These sequences can be cloned
XX      into plasmids and used to transform E. coli to produce the DCP
XX      proteins. DCP peptides comprise repeated units of: A = GAPGAGP,
XX      B = GSRGDPGP and/or C = GAHGAPGPK. These polymers may be used to
XX      raise anti-DCP antibodies in rabbits. These polymers have molecular
XX      weights of >30 kD and are able to form helices due to interchain
XX      linkages. These polymers pref. contain a proportion of tripeptide
XX      triad sequences found in natural collagens, pref. mammalian collagens.
XX      The CLPs impart unique characteristics to materials such as fibres,
XX      membranes, films, coatings, hydrogels, colloid suspensions and moulded
XX      articles.
SQ      Sequence 756 BP; 84 A; 288 C; 291 G; 93 T; 0 other;
```

Query Match 6.7%; Score 49; DB 14; Length 756;
Best Local Similarity 43.5%; Pred. No. 0.27;
Matches 223; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

```
QY      72 gtcggaagcttggaagcgcgcgcatgtgcccggggcgctaccgggtgatacaaacgt 131
DB      218 gtgcgcttggaagcgcgcgcatgtgcccggggcgctaccgggtgatacaaacgt 277
QY      132 atggggcgaggagaccccgatgagtcggaacttgaaagggcaactcaacgat 191
DB      278 ctggaccgctgtgtccacgggtgctccgggaacttgaaagggcccgccaggtgcccggac 337
QY      192 caacggcgatcacagcaacggcaacagctgcccgtaccctaccgtactactcgg 251
DB      338 cggctgtgtccacgggtgtgctccgggaacttgcaagcccgacaggtgcccggacgctg 397
QY      252 gtccactggggcgcttgcaaggaacttcggatgcccggcgcttgcaaggaactgtc 311
```


PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis

PS Claim 4; Page 187-188; 250pp; English.

CC This is the 3' region of DNA coding for an antigenic portion of
CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in
CC AA44438.. XP25 DNA was isolated from a M. tuberculosis strain Erdman
CC genomic DNA expression library using sera from patients having
CC extrapulmonary tuberculosis. It bears no similarity to known
CC sequences. The invention relates to methods and compositions for
CC diagnosing tuberculosis. It provides polypeptides (see
CC AAM64291-W64379) comprising an antigenic portion of a soluble M.
CC tuberculosis antigen, or an immunogenic portion of a M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic
CC kits for detecting M. tuberculosis infection in a patient using
CC these polypeptides, antibodies or oligonucleotide probes and
CC primers.

SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.7%; Score 48.8; DB 19; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.3;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 13 ttcccgatgagcgaacgaagaacgagcctgagcccgagccgagcgtgagctg 72
DB 175 ttcccgatgagcgaacgaagaacgagcctgagcccgagccgagcgtgagctg 234
OY 73 tgcgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 132
DB 235 accggtacgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 294
OY 133 tgggagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 192
DB 295 gccgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 354
OY 193 acaagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 252
DB 355 ggccttgaagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 414
OY 253 tgcgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 308
DB 415 atcaagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 474
OY 309 gtcgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 368
DB 475 gccgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 533
OY 369 ctacgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 428
DB 534 cggagagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 593
OY 429 gctgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 488
DB 594 gggagagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 653
OY 489 cgtgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 548
DB 654 cggagagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 713
OY 549 cgcctacgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 602
DB 714 cactgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 767

RESULT 11
AA219349
ID AA219349 standard; DNA: 985 BP.

XX AA219349;
AC
XX
XX 05-NOV-1999 (first entry)
DE
XX
XX M. tuberculosis antigen 3' XP25 DNA sequence.

KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KM immunotherapy; diagnosis; immunisation; vaccine; infection;
XX immune response; skin test; ss.

OS Mycobacterium tuberculosis.

PN W09942076-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99MO-US03268.

PR 05-MAY-1998; 98US-0072967.

PR 18-FEB-1998; 98US-0025197.

PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;

DR WPI; 1999-527409/44.

PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions

PS Claim 4; Page 171-172; 299pp; English.

CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
CC AA219225 are used in the exemplification of the present invention.

SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.7%; Score 48.8; DB 20; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.3;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 13 ttcccgatgagcgaacgaagaacgagcctgagcccgagccgagcgtgagctg 72
DB 175 ttcccgatgagcgaacgaagaacgagcctgagcccgagccgagcgtgagctg 234
OY 73 tgcgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 132
DB 235 accggtacgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 294
OY 133 tgggagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 192
DB 295 gccgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 354
OY 193 acaagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 252
DB 355 ggccttgaagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 414
OY 253 tgcgagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 308
DB 415 atcaagcgtgagcgaacgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 474

FT	/product= "ORF5"
FT	complement (5392..6147)
CDS	
FT	/tag- h
FT	/product= "ORF6"
RBS	complement (6152..6156)
FT	/tag- i
CDS	complement (6194..7282)
FT	/tag- j
FT	/product= "ORF7"
CDS	complement (7280..8133)
FT	/tag- k
FT	/product= "ORF8"
RBS	complement (8141..8145)
FT	/tag- l
FT	complement (8254..9318)
CDS	
FT	/tag- m
FT	/product= "ORF9"
RBS	complement (9324..9328)
FT	/tag- n
FT	9575..10504
CDS	
FT	/tag- o
FT	/product= "ORF10"
RBS	9568..9571
FT	/tag- p
FT	10584..11585
CDS	
FT	/tag- q
FT	/product= "ORF11"
PN	
PN	MO200151639-A2.
XX	
PD	19-JUL-2001.
PF	
PF	12-JAN-2001; 2001WO-US01187.
XX	
PR	12-JAN-2000; 2000US-0175751.
XX	
PA	(SCHE) SCHERING CORP.
PI	
PI	Hosted TV, Horan AC, Wang TX;
XX	
DR	WPI: 2001-442147/47.
DR	P-RSDB: AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906,
AAU04907, AAU04908, AAU04909, AAU04910, AAU04911.	
XX	
PT	New nucleic acid molecules encoding evernimicin pathway gene
PT	products, useful for improving yields of evernimicin, to produce new
PT	evernimicin and as probes to identify homologous sequences -
PS	
XX	
XX	Example 1; Fig 12; 109pp; English.
CC	The sequence encodes 11 proteins comprising enzymes of the
CC	evernimicin antibiotic biosynthetic pathway. A vector comprising a
CC	M. carbonacea evernimicin biosynthetic pathway resistance gene product
CC	is useful for selecting for a transfected or transformed host cell. An
CC	integrative version of the vector is useful for introducing a
CC	evernimicin pathway gene (a bottle-neck gene) into an actinomycete of
CC	the genus Micromonospora. The DNA encoding the biosynthetic proteins is
CC	useful for synthesising novel evernimicin-related compounds, arising
CC	from modifications of the DNA sequence designed to change glycosyl and
CC	modified orsellinic acid groups contained in evernimicin, for
CC	expressing functional or mutant evernimicin biosynthetic enzyme for
CC	evaluation, diagnosis and preferably biosynthesis of evernimicin or
CC	other secondary metabolic products, improving the yield of evernimicins
CC	and to produce novel evernimicins and also as a hybridisation probe to
CC	identify homologous sequences. The encoded polypeptides are useful for
CC	combinatorial biosynthesis to generate libraries of orthomycins, e.g.
CC	evernimicin analogues/homologues and drug discovery. The
CC	DNA encoding the integrase allows for increasing a given gene dosage. The
CC	integrative vector can be used to permanently integrate copies of a
CC	heterologous gene of choice into chromosomes of different hosts and to
CC	integrate genes which increase the yield of known products or to generate
CC	novel products such as hybrid antibiotics or other novel secondary
CC	metabolites. The vector can also be used to integrate antibiotic
CC	

CC resistance genes in order to carry out bioconversions with compounds to
CC which the strain is normally sensitive and is thus useful in fermentation
CC processes involving e.g. *Streptomyces* antibiotics.
XX
SQ Sequence 12152 BP; 1675 A; 4502 C; 4475 G; 1500 T; 0 other;

Query Match 6.6%; Score 48.4; DB 22; Length 12152;
Best Local Similarity 45.9%; Pred. No. 0.4;
Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

[illegible]

```

RESULT 14
AAI99683/C
ID AAI99683 standard; DNA; 4403765 BP.
XX
XX
AC AAI99683;
XX
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SFG ID NO. 2.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX OS Mycobacterium tuberculosis.
XX
XX US6294328-B1.
XX
XX PD 25-SEP-2001.
XX
XX PF 24-JUN-1998; 98US-0103840.
XX
XX PR 24-JUN-1998; 98US-0103840.
XX
XX PA (GENO-) INST GENOMIC RES.
XX
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX DR WPI; 2001-647261/74.
XX
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ

```

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:45:36 ; Search time 130.83 Seconds

(without alignments)
1374.331 Million cell updates/sec

Title: US-10-003-759-3_COPY_52_783

Perfect score: 732
Sequence: 1 tgcgactgcgtcttcctccga.....atttccgtaacgltgcag 732

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
- 4: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
- 5: /cgn2_6/prodata/2/lna/PCBUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/lna/Backfileseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.8	18.0	1116	US-09-104-308-2	Sequence 2, Appl1
2	131.8	18.0	1116	US-09-321-981-2	Sequence 2, Appl1
3	105.8	14.5	1716	US-09-321-981-4	Sequence 4, Appl1
4	49	6.7	756	US-08-642-255-50	Sequence 50, Appl1
5	48.8	6.7	985	US-09-056-556-182	Sequence 182, Appl1
6	48	6.6	4403765	US-09-103-840A-2	Sequence 2, Appl1
7	47.6	6.5	30001	US-08-125-468-1	Sequence 1, Appl1
8	47.6	6.5	30001	US-08-474-933-1	Sequence 1, Appl1
9	47.2	6.4	1908	US-08-173-508-1	Sequence 1, Appl1
10	47.2	6.4	1908	US-08-265-310-1	Sequence 1, Appl1
11	47.2	6.4	1908	US-08-951-742-1	Sequence 1, Appl1
12	46.8	6.4	15872	US-09-105-537-1	Sequence 1, Appl1
13	46.2	6.3	1288	US-08-440-856A-9	Sequence 9, Appl1
14	46	6.3	44377	US-08-804-227C-7	Sequence 7, Appl1
15	46	6.3	44377	US-08-804-198-1	Sequence 1, Appl1
16	45.2	6.2	1291	US-08-997-897-1	Sequence 1, Appl1
17	45.2	6.2	1291	US-09-156-836B-1	Sequence 1, Appl1
18	44.6	6.2	2852	US-09-056-556-203	Sequence 203, Appl1
19	44.6	6.1	1176	US-08-387-942C-17	Sequence 17, Appl1
20	44.6	6.1	11230	US-09-105-537-32	Sequence 32, Appl1
21	44.6	6.1	12588	US-08-387-942C-1	Sequence 1, Appl1
22	44.6	6.1	36778	US-09-105-537-5	Sequence 5, Appl1
23	44.6	6.1	38506	US-09-320-878-19	Sequence 19, Appl1
24	44.6	6.1	2712	US-09-025-691-4	Sequence 4, Appl1
25	43.6	6.0	1017	US-08-849-751-1	Sequence 1, Appl1
26	43.6	6.0	1017	US-09-478-816-1	Sequence 1, Appl1
27	43.4	5.9	1500	US-09-593-711A-10	Sequence 10, Appl1

28	43	5.9	1273	4	US-09-319-892-3	Sequence 3, Appl1
29	43	5.9	1333	4	US-09-372-422A-9	Sequence 9, Appl1
30	42.8	5.9	599	1	US-08-584-226-3	Sequence 3, Appl1
31	42.4	5.8	20235	3	US-07-642-734C-3	Sequence 3, Appl1
32	42.4	5.8	20235	3	US-08-439-009A-3	Sequence 3, Appl1
33	42.2	5.8	1158	4	US-09-372-422A-21	Sequence 21, Appl1
34	42	5.7	71989	4	US-09-443-501A-2	Sequence 2, Appl1
35	41.8	5.7	432	1	US-08-642-255-48	Sequence 48, Appl1
36	41.6	5.7	1155	2	US-08-913-264-2	Sequence 2, Appl1
37	41.4	5.7	669	3	US-08-387-942C-7	Sequence 7, Appl1
38	41.4	5.7	18994	1	US-08-459-586-4	Sequence 4, Appl1
39	41.4	5.7	18994	2	US-08-282-696-4	Sequence 4, Appl1
40	41	5.6	1182	4	US-09-385-028-19	Sequence 19, Appl1
41	41	5.6	8051	2	US-08-576-626A-2	Sequence 2, Appl1
42	41	5.6	11604	4	US-09-385-028-13	Sequence 13, Appl1
43	41	5.6	15079	4	US-09-385-028-1	Sequence 1, Appl1
44	41	5.6	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
45	41	5.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-104-308-2
; Sequence 2, Application US/09104308
; Patent No. 6187577
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: Van Der Kleij, Wilhelmus A.H.
; APPLICANT: Weyler, Walter
; TITLE OF INVENTION: No. 6187577e1 Cellulase Producing Actinomycetes,
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genecor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,308
; FILING DATE: 24-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/974,042
; FILING DATE: 19-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC539
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7555
; TELEFAX: 650-846-6504
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-104-308-2
Query Match 18.0%; Score 131.8; DB 4; Length 1116;
Best Local Similarity 50.7%; Pred. No. 2e-19;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

```

1  Oy      47  ctyagagccgagccgagccgctctgtcggaacgctgaggacgcgcgatgtgtgcccggg  106
2  Db      77  CGGGCTCCGCCCCAGCGGAACCAAGATCTGGACCGGCTTAAGCAGCACCAACATCAGG  136
3  Oy      107  ggcgcataccgggtgtatcaacaacgtatgggcgcgaggagacgcgcagtgatctgagtcg  166
4  Db      137  ACCGGTATCTGGTGCAGAACCAACCCCTGGGGCCACGACGCGACCAACCCATGCATATGTGA  196
5  Oy      167  gacttggaaacggcgactctacgatlacacacgggccgatacggacatcggaaacggaacagtg  226
6  Db      197  CCGGCACACGGTTTCGAGATCACCCAGGCCGACGCTTCGGTCCGACCAACAGGGCCCCGA  256
7  Oy      227  ccgcctatccggccactctactctcggtgtgcgaactggggcgcccttcaacgagcaattccggat  286
8  Db      257  ACTCCATATCCCTCGGCTCAAGACAGCGCTGCACACTAACGCAACTCGGGCCCCCGACAGCCG  316
9  Oy      287  tgcgcgcyggcgctgtcagagagctgtccgaacgttgcgaacgagcttgyaagcttaccgcgata  346
10 Db      317  TCCCCATGCGGATCAAGCTGATCGGAGGCGCCGCCACAGATGTCTCTACCGCTACACCG  376
11 Oy      347  cgaaggccgctggaatgtgcgcctacgacatacttgttctagtcgccgtcacaagaattccgga  406
12 Db      377  GGAAGGGCTTATACAAGCGCGGTACGACATCTGG-----CTGGAACCGCACACCCCGCA  430
13 Oy      407  acggtctacagcgcgcgagctgtgatacttctgtgcctgaacttggaaacggcgcgctgtatgc  466
14 Db      431  CCAAGCGGGGTGAACCGGACCGAGATCATCTGTTCACACCGGTGCGGCCGCTCAGC  490
15 Oy      467  cggcgcgagccgctgtgcacccgtggaacttggcggggcacacttggaaacttggatgtgtatg  526
16 Db      491  CCATCGCTTCCGCGGCTGCGGACGCGCCACGTCGCGCGGCCCGCCAGCTGGAGGTGTGGACCG  550
17 Oy      527  ccgaactgggaactgtgaattacatctgcctacgcgcgacagcgcacccaacgltcgtgtgagcg  586
18 Db      551  GCAGCAACCGTTTCGAACGACGATCTCCTTCCTGCGCGCCCTCCGGATACGACAGCTGA  610
19 Oy      587  agcttgaaccttgaagcgctcatctcgaagaacgggtgtgcgcggcgctacatctccgcccggagt  646
20 Db      611  GCTTGACGCTCAAGGACTTCTGTCGACCAAGCCCTCAAGCCAGCGGCTGGGCTACCCCGGACT  670
21 Oy      647  ggtatctgacatgtcgtgtggaacgggcttcgaaactcttggaaaggagggcgggcgctgtcgaa  706
22 Db      671  GSTACTCTCACGACATCCAGCGGCGCTTCGAGCCGTGGAGCGGCGGACCGGACGCTGTGGCGG  730
23 Oy      707  ggcgcgcatlcttcgtaacggtgca  731
24 Db      731  TGAACCTCGTTCTCCGCGGTGAA  755
25
26 RESULT 2
27 US-09-321-981-2
28 : Sequence 2, Application US/09321981
29 : Patent No. 6287839
30 : GENERAL INFORMATION:
31 : APPLICANT: Genencor International, Inc.
32 : TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomyces,
33 : FILE REFERENCE: G0540-2
34 : CURRENT APPLICATION NUMBER: US/09/321,981
35 : CURRENT FILING DATE: 1999-05-28
36 : PRIOR APPLICATION NUMBER: PCT/US99/11971
37 : PRIOR FILING DATE: 1999-05-28
38 : PRIOR APPLICATION NUMBER: 09/104,308
39 : PRIOR FILING DATE: 1998-06-24
40 : PRIOR APPLICATION NUMBER: 08/974,042
41 : PRIOR FILING DATE: 1997-11-19
42 : NUMBER OF SEQ ID NOS: 5
43 : SOFTWARE: FastSeq for Windows Version 3.0
44 : SEQ ID NO 2
45 : LENGTH: 1116
46 : TYPE: DNA
47 : ORGANISM: Unknown
48 : FEATURE:
49

```

```

; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
; OTHER INFORMATION: thermovibrio
US-09-321-981-2

```

Query Match	18.0%;	Score 131.8;	DB 4;	Length 116;
Best Local Similarity	50.7%;	Pred. No. 2e-19;		
Matches 347;	Conservative 0;	Mismatches 332;	Indels 6;	Gaps 1

OY	47	ctgagccgagcgccgacgctgagagctctgagagacgctggagagacgagcgagcttgagccggg	106
Db	77	cggtctccgcgcacgagcgcaaacccagagatctgcgacgcgtacagacccacccagatccag	136
OY	107	gagcctacccggtgtatccaacagatctgagggcgcgagagacccgaccagtgatctgaggtc	166
Db	137	accggtacgtggtgtcagagaaacacccgctgggagccacgagccacccagtgatccaatgga	196
OY	167	gactcgagaaacgggccaacttcagatccaacccggccgcatccaagacaagcgcaaacggtg	226
Db	197	ccggcaaacggtcttcgagatccaacccagagcgggtctcggtgcgacccaaagcgcccca	256
OY	227	ccgcgtactccggccacatctacttcgggttgccacatctgggctgcgcctccagacagacaa	286
Db	257	agtcctactccctcgtgtctacgagcggtctgcacactacgacaaatcgtcgcccccagacga	316
OY	287	tgcccgcgcgcgctgacagagctgtccggaagctgcgcacgagcgtcgagcgtccaagccga	346
Db	317	tgcccaacgctgagatcagatcgtctcgacagcgcgcccaagcagtgctctctctacccgta	376
OY	347	cgagcgagccgctggaatctgcgcctacagacatctggtctcagtcgcgtctaaagaa	406
Db	377	gcaacggcgctctacaacacgcgcgctgacgacatctgg-----ctgaaaccgaaacccgca	430
OY	407	acgggtcacagccggcgagcgcgcgagctgatatctgtgcctggaacttgaaacggcggtgatgc	466
Db	431	ccaaacgggggtggaacccggaaacggagatcatgatctgtgtctcaacccgggtcgccgggtccgc	490
OY	467	cgagcgagccagccgctgagccacggctgagaaactgcccgggcccacacctggagaaagtctgata	526
Db	491	ccatcgggtctgcgcgtctgcgcgcgcccaagctcggcgcgccgcgacgtggagagtggtggacgg	550
OY	527	ccgagctggaactggaattacatctgcctacccggcgccgcagacccacacagctcgtgtgaacg	586
Db	551	gcagcaaacggttcgaaagacgtgatatctcctctcctggcgccctcggatccaagcaagctgga	610
OY	587	agctcgagccctgaaagcctctcatccagacgacgggtcgcccgcgctcaacatccggcccgaggt	646
Db	611	gcttcgcagcgtcgaagaaatctgtcgaacagagccgtccagcgcccttggccaccccgagact	670
OY	647	ggtatctgcgaatcggtgtgagacgggctctcgaaactctctggagagcggggcgagctcgtaa	706
Db	671	ggtatctccacacagatccacgagcggtcttcgagacggttggaagggcggaacccggtctggcg	730
OY	707	gcgcggaattttccgtaacgggtgca	731
Db	731	tgaactcgtctctcctccgcggtgaa	755
RESULT 3			
US-09-321-981-4			
Sequence 4, Application US/09321981			
Patent No. 6287839			
GENERAL INFORMATION:			
APPLICANT: Genencor International, Inc.			
TITLE OR INVENTION: No. 6287839e1 Cellulase Producing Actinomyces,			
FILE REFERENCE: GCS40-2			
CURRENT APPLICATION NUMBER: US/09/321,981			
PRIOR FILING DATE: 1999-05-28			
PRIOR APPLICATION NUMBER: PCT/US99/11971			
PRIOR FILING DATE: 1999-05-28			
PRIOR APPLICATION NUMBER: 09/104,308			
PRIOR FILING DATE: 1998-06-24			


```

RESULT 7
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantioli, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/125,468
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEO ID NO.: 1 :
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match          6.5%; Score 47.6; DB 1: Length 30001,
Best Local Similarity 45.3%; Pred. No. 0.076;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

OY 144 gaccgccagtcgattgagtgcgaactggaaacggcgcaactcagatcacagcgccga 203
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25022 GACCGGCAAGGGCTTCGCCCTCGCGGCGAGGGCGCGCGGATGTTCGTCTGAGAGCGGGA 24963

OY 204 tcagacaacaggcaaacacagctggccgcatatccggccatctacttggtgtgccactggg 263
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24962 GCACCGCAGTCCCGCGCGCGGCCACACTTACGCCGCGAGGTCCACCGGCTACGCCCGGCGCTG 24903

OY 264 gccctcgacacgaacttgcggatctgcggcgcgcggtgcagggagctgtagcagctggcgac 323
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24902 CACTCTCTTACACATGATCAGCGGCGCTGGGGTTGCGACGGGGAACGAGATGCGCACGCTATCCG 24843

OY 324 gagctgagcgtccaacgcgatcatcagacagcgcgctgtgaaatgcgcgtatacagacatgtgt 383
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24842 GGTC---GGCGTCGACCGGAGCGCGGATGAAACCCCGAGGACGTTGGCTACCTCAACCGGCA 24786

OY 384 cagtcgcgttacgaattccggaacggctacagcgcgcgcgcgagcgcgagctgatatgttgt 443
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24785 CGGCTCGGCGACCAA---GCAAGACGACCGGACGACGAGACCGCGCGCTTCAAGAAGAGCT 24729

OY 444 gaactcgagaacggcggtgatatgctcgcgcgcgcgacgcgcgctgtgacacgctggaactggcg 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24728 GGGCGACCAACGCTTACCGGGTGCGGGTCAAGCTCATGATCAAGTCAATGATGAGAACCTCGCT 24669

```

[illegible]

Db 25022 GACCCGACAGGCTTCGCTCGGCGAGGCGCGCGATGTTGCTGGAGAGCCGGA 24963
QY 204 tcacgaacagcgacaacagctggccgcctatccgagccatctactctggcgcaactgg 263
Db 24962 GCACGCGATCGGCGCGCGCGACATCTACGCCGAGGTACC GGCTACGCCGCGCTG 24903
QY 264 cgcctcgacaggaatcggagatcgcgcgcgctgcaagagctgtcgcagctggcgac 323
Db 24902 CAACTCTACAGATGACCGGCTGGCGGTGACAGGGACAGATGGCCAGGCTATCCG 24843
QY 324 gaactggaagctcaagcgagctcaagagcgcgctggaatgcgcgctcagacatctg 383
Db 24842 GGTG---GCGCTGACCGCGACCGGATGAAACCCGAGACGTGCGCTCAACCGCA 24786
QY 384 caatccgctcaagatcgcgaagcgctacagcgcgcgagctgagctgagctgctgct 443
Db 24785 CGGCTCGGCGACCA---GCAGAACGACCGGACAGACCGCGCGCTTCAAGAGACCT 24729
QY 444 gaactggaacgcgcgctgagctgagcgcgcgcgctgagcgcgctgagcgcgcgcg 503
Db 24728 GCGCGACACAGCGCTACCGGCTGCGGTGACGCTGATCAATGATGATCGACACTGCT 24669
QY 504 ggcacactgggaagctgctgagctgagcgcgcgctgagctgagctgagctgagctgag 563
Db 24668 GGGCGCGCTGCGCTGAGTTCCGCGCGACCGCGCTGCGCTGAGACACTGCGTGT 24609
QY 564 gacgcacacacgctgctgagctgagctgagcgcgctgagcgcgctgagcgcgctgag 623
Db 24608 GCGCGCGACCGCGCTGAGTTCCGCGCGACCGCGCTGAGACACTGCGTGTGCGCT 24549
QY 624 cgcgagctacatcgcgagctgagctgagctgagctgagctgagctgagctgagctg 683
Db 24548 GGTGGCGCGGAGACAGCGGTGCGGACAGGTGCTCAGCGTGGCGGCTTCCGCGCT 24489
QY 684 ggaagcgcgcgcgctgct 701
Db 24488 CCAGACGCGATGCTGT 24471

RESULT 9
US-08-173-508-1
; Sequence 1, Application US/08173508
; Patent No. 5616485
; GENERAL INFORMATION:
; APPLICANT: Bartfield, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,508
; FILING DATE: 23-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/125/CACO
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1908 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 146..1759
FEATURE:
NAME/KEY: misc-feature
LOCATION: 146..148
OTHER INFORMATION: /note="Met at position -39
represents fmet"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 146..262
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 263..1756
US-08-173-508-1

Query Match 6.4%; Score 47.2; DB 1; Length 1908;
Best local Similarity 46.7%; Pred. No. 0.076;
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

QY 320 gcaagagctgagcgtcaagcgagctacagagcgcgctggaatgcgcgctcagacatc 379
Db 159 GCATACGGCGGAGGCGGACCGCTTCGCGACGCGCGAGACACTGTCACCGCGACGCTGA 218
QY 380 ggtcagctccgctcaggaatcgcgcaagcgctacagcgcgcgcgctgagctgagct 439
Db 219 TCGCCGCGCGCTGCTGCGCACCCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
QY 440 ggtcagctgagcgtcagcgcgctgagctgagcgcgcgcgctgagcgcgcgctgagcg 499
Db 279 GCGGAGAGCTGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
QY 500 cgcggcgacactgagctgagctgagctgagctgagctgagctgagctgagctgagct 556
Db 339 CGGCGATCGACTGGAGGAGCTGCGGACCGGACGGAACCTGCCCAAGCCCATCTCACTGCG 398
QY 557 ggcgcacagcgccacacagctgagctgagcgagctgagcgcgcgctgagcgcgctcag 616
Db 399 GCTACGTACAGGTGCGGATGAGCTACGCCAAGCCGTAAGGCAAGATCAGGCTGCGCG 458
QY 617 cgttcgcccgcgctacatccgcgcgagctgagctgagctgagctgagctgagctgagct 676
Db 459 TCGACCGCATCGGCAACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 518
QY 677 aactcggagagcgcgcgcgctgagcgcgcgctgagcgcgcgctgagcgcgcgctg 712
Db 519 CCGCGGCTCCGCGCGGCTCGCGCTGCGCTGCTTCCCGG 554

RESULT 10
US-08-265-310-1
; Sequence 1, Application US/08265310
; Patent No. 5856166
; GENERAL INFORMATION:
; APPLICANT: Bartfield, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; APPLICANT: Malek, Lawrence T.
; APPLICANT: Soostmeyer, Gisela
; APPLICANT: Walczyk, Eva
; APPLICANT: Kryzman, Phyllis

```

APPLICANT: Garven, Sheila
TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W.
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,310
FILING DATE: 24-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,508
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 18740/133/CACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 672 5300
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1908 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 146..1759
FEATURE:
NAME/KEY: misc.feature
LOCATION: 146..148
OTHER INFORMATION: /note="Met at position -39
OTHER INFORMATION: represents fMet."
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 146..262
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 263..1756
US-08-265-310-1
Query Match 6.4%: Score 47.2; DB 2; Length 1908;
Best Local Similarity 46.7%: Pred. No. 0.076;
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1.
320 gcaagagctgtgacgtctacgcgcgtacgcaagcgagcgctgtgatatcgcgctctcgaatct 379
Db 159 GCATACGGCGGAGGGCGCGCGCTTCGGACGCGCGGAGACACTGTGTACCGACGCTTA 218
Qy 380 ggttcagctcccgctacgcatctccggcaacggttacagcgcgcgccgagctgatcatct 439
Db 219 TCGCGCGGCGCGGTCTCGGCGACCGCGCGGAGCGCGCGCCCGCGGCGACGGCGGCGACG 278
Qy 440 ggtctgaactgtgaacggcgcgctgatatgccggcgcgagccgcgttgccacacgtggaactgg 499
Db 279 GCGCGAGCTGGGACCGCGGAGCGCGCGGTGCGCGCATCGCGCGCGCGCGCGCGCGCGG 338
Qy 500 ccggggccacctgtggaagtctgtgtatgcccagctgtgagctgacgctgacacatcgctacc 556
Db 339 CGGGCATCGACTGCGAGGAGCTGCGCGAGCGCGGATGTGAACCTGCGCCCAAGCCCATTCAGTGGC 398

```

```

Oy 557 ggcgcagcagcccccaccacgtcgttgcagcagctgcagcctcgaagcctcatgcagcagc 616
      | 1111111111111111111111111111111111111111111111111111111
Db 399 cctacgttcacgcgtcccatgcactgcactcaccacaaccctgacgcacacagatcagctccgc 458
      | 1111111111111111111111111111111111111111111111111111111
Oy 617 cgtcgcgcgcgcgcgtacatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 766
      | 1111111111111111111111111111111111111111111111111111111
Db 459 tcgaccgcgcatgcgcacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 518
      | 1111111111111111111111111111111111111111111111111111111
Oy 677 aactctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 712
      | 1111111111111111111111111111111111111111111111111111111
Db 519 ccggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 554
      | 1111111111111111111111111111111111111111111111111111111

RESULT 11
US-08-951-742-1
; Sequence 1, Application US/08951742
; Patent No. 6127144
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Michael J. Butler
; APPLICANT: Dany Hadary
; APPLICANT: David Jenish
; APPLICANT: Tim Krieger
; APPLICANT: Lawrence T. Malek
; APPLICANT: Gisela Soostmeyer
; APPLICANT: Eva Walczyk
; APPLICANT: Phyllis Krygsman
; APPLICANT: Sheila Garven
; TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN
; TITLE OF INVENTION: BACTERIAL HOST CELLS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951.742
; FILING DATE: 16-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 0189740/0140
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1908 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..1756
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 146..148
; OTHER INFORMATION: /product= "Met at position -39
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 146..262

```



```

Db 459 GCGCGTGTGGCGSNRYGRGSGSCGS---CTGACGCTGYWSTGCAACACGCCGGGGTGGT 515
OY 229 gccatccggccatctactctcgggttcacattgagcgccgacaggaacattgagattg 288
Db 516 GCGCGCCAGAGCGCCGGGCCAKAGCATCTGTGCTTTCACACGSGSSGAGTTGACGCC 575
OY 289 ccgagcgcgctgcagagagctgtccagacgttcgcagacagactgcgaactcagccgatacag 348
Db 576 GTGCTCCGCGTCAACGCGGCTGTGGGGCGCGCTCGGGSATGAAAGCACCGGGCGCGCCCATG 635
OY 349 acgagcgcgctggaatgcgcgcctacgaacatttgttcaagtcctcgatcaagaattccggaac 408
Db 636 ACCGAGGCGCGCGCGG-CGGGGGAGGAGTCRTCTCCCTCCACACCGCTCCGSGCGTGGCTTS 694
OY 409 ggcacacagcgcgcgcgccgagactgatctgacgtgaactgcgaagcgcgcgctgatgcg 468
Db 695 GCGCGGCTCGGCGCCGCGCACGCTTACACCGGCTCCCAAGCACCGCATGTGSGGGCTCACCAAG 754
OY 469 ggcgcagcgcgctgtgcccacccgttgcgaactgacgagcgcgggccacactgcggaaagctgtatgcc 528
Db 755 AACCGCCGCTCGAGCGTSSGGGCGSCACGSGSRTCCGSGTCAACTCCTTCCTCCCTTGCGC 814
OY 529 gactggagacttgaattacattcattcgtctaccgagcgacgcgcgccacacaaagtcgtgagagag 588
Db 815 GTCCGCAACSCCSATGCTCATTAACGCTGTGGCGCCAGGGCCACGACGAKCKCACCGCCGAC 874
OY 589 ctgagactgaagagcttcattcatgcagacgagctgcgcccgcggtctacacccgagcgagtg 648
Db 875 GMGACCCGAGACSTCGACMTCGACTCTGCGACTCTGCCCTGCCACGACGACGAGAGCTTGAC 934
OY 649 tatctcaatcggttgcgaagcgacttcgaaactctgcgaagggcgagggc 695
Db 935 AAGATGGAGGAGTGTCGTGACGAGGCGCTCGCACGCTCAAGGAGGCGSCSAC 981

RESULT 14
US-08-804-227C-7/c
? Sequence 7, Application US/08804227C
? Patent No. 5876991
? GENERAL INFORMATION:
? APPLICANT: Dehoff, Bradley S.
? APPLICANT: Kuhstoss, Stuart A.
? APPLICANT: Rostock, Paul R., Jr.
? APPLICANT: Sutton, Kimberly L.
? TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
? NUMBER OF SEQUENCES: 15
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: THOMAS G. PLANT 1501
? STREET: LILLY CORPORATE CENTER
? CITY: INDIANAPOLIS
? STATE: IN
? COUNTRY: USA
? ZIP: 46285
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: ASCII(DOS) Text only
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/804,227C
? FILING DATE: February 21, 1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Plant, Thomas, G.
? REGISTRATION NUMBER: 35,784
? REFERENCE/DOCKET NUMBER: X-8231
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 317-276-2459
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 44377 base pairs
? type: nucleic acid
?

```

```

: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 350..14002
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 14046..20036
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 20110..31284
:   NAME/KEY: CDS
:   LOCATION: 31329..36071
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 36155..41830
: US-08-804-227C-7
:
: Query Match 5.3%; Score 46; DB 2; Length 44377;
: Best Local Similarity 47.6%; Pred. No. 0.17;
: Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
:
: Oy 19 gatgcgcaacaacggaagaacccggaagccttgaagcccgagccgacccgtgagctgtgcgga 78
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 18968 GAGCGCGCGCGTCGCGCTTGCGGCGCGGAGCACCCTGTGCATCCGCTCGGCGGCGGGA 18909
:
: Oy 79 cgtctggagcgcgcggaatgttgccgggggggcgcctacgcgggtatcaacaacgatggggc 138
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 18908 GAGCAGCGCGTCGTGTGACACACCCCGCGCGGTGTGCACACGACAGTCACGGGCTTTCTGTC 18849
:
: Oy 139 gcggagaccgcgcgaatgtatgtaggttcggaactggaacgggcaacttcagatcacagc 198
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 18848 GGGGAGGTGCGTGTGACAGGCGCGCGCGCGGTGTGCACACGACAGTCACAGGCCCGTAC 18789
:
: Oy 199 gccgatacgaacaacggcaacagtcgtgcgcctatccggccattacttgggtggtgcac 258
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 18788 GCGCGACTTCGGACCGACCGACTCGTCACTCGCGACCAATGCATGACGCCCGGTGCGTC 18729
:
: Oy 259 tggggcgctctgacgagcaatctcggaattgcgcggcgcgcgatgcagg 304
:   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
: Db 18728 GGGACCGCGCCCGCCACCACGACGAGATGGCGCACCTTGTGCCG 18683
:
: RESULT 15
: US-08-804-198-1/c
: Sequence 1, Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
:   APPLICANT: Burgett, Stanley G.
:   APPLICANT: Kuhstoss, Stuart A.
:   APPLICANT: Rao, Nagaraja R.
:   APPLICANT: Richardson, Mark A.
:   APPLICANT: Rosteck, Paul R., Jr.
:   TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
:   NUMBER OF SEQUENCES: 6
:   CORRESPONDENCE ADDRESSES:
:     ADDRESSEE: PAUL R. CANTRELL 1138
:     STREET: LILLY CORPORATE CENTER
:     CITY: INDIANAPOLIS
:     STATE: IN
:     COUNTRY: USA
:     ZIP: 46285
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: Macintosh
:   OPERATING SYSTEM: Macintosh 7.0
:   SOFTWARE: Microsoft Word 5.1
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/804.198
:   FILING DATE:
:

```

```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470
REFERENCE/DOCKET NUMBER: P9113
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-3885
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-08-804-198-1
```

```
Query Match 6.3%; Score 46; DB 2; Length 44377;
Best Local Similarity 47.6%; Pred. No. 0.17;
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 19 gatggcgaaacggaagaaagcctgagcccgacccgagcgtgagctgagctgagcga 78
Db 18968 GACGGCGGCGTGGCGCTTGGCGCGGACACCGTGTGATCCGCTCGCGCGTCTGGCGGA 18909

QY 79 cgcttggaacgcgcgagatggtgcccgggggagctacacgggtgatcaacaagatgagggc 138
Db 18908 GAGCACGGCGTGTGTGACGACACCCGCGGTGTGACACGACGATCAGCGGCTTTCGTC 18849

QY 139 gcggaacgcgcgcagtcgagtcggaactggaacgaggaactcaagatcaacagc 198
Db 18848 GGGGAGTGTCTGACAGCGCGCGCGCGGTCTGCGGTGACGATCAGGCGCGGTAC 18789

QY 199 gccggtcacggaacggaacaaagtcgagcgtatccgacatctacttctggtgacac 258
Db 18788 GCGGACCTCGGACCGCGACTCGGTCACTCGGCGACAGTCGATGACCGCGGTGCGTC 18729

QY 259 tggggcgagcctgcaaggaatcgggagtcgagtcgagcgcgagtcgagc 304
Db 18728 GGGACCGCGCGCTGCCACACAGAGAGATGGCGCACCTTGTCCCG 18683
```

Search completed: August 19, 2002, 14:01:35
Job time: 14899 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 10:59:53 : Search time 3821.57 Seconds
(without alignments)
2585.265 Million cell updates/sec

Title: US-10-003-759-3_COPY_52_783

Perfect score: 732
Sequence: 1 tgcgacgcgcctctccga.....atttcgcgtaacggtgcag 732

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_lin:*
15: em_gss_pin:*
16: em_gss_vrti:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	60.8	8.3	925	12	CNS0091P
2	56.2	7.7	925	12	CNS0091P
3	55.6	7.6	932	12	CNS00720
4	54.4	7.4	864	10	BF253875
5	53.8	7.3	964	12	CNS003WG
6	50.6	6.9	700	12	AG060149
7	50.4	6.9	935	12	CNS006XK
8	50.4	6.9	1022	9	BE040131
9	50	6.8	894	12	AG060214
C 10	49.8	6.8	1009	12	CNS010EW
11	49.6	6.8	449	10	BE017742
C 12	49.2	6.7	534	12	AZ934730
C 13	49.2	6.7	534	12	AZ934730
C 14	49.2	6.7	546	12	AZ934730
C 15	48.2	6.6	443	9	AU030637
16	48.2	6.6	443	9	AU030637
17	48	6.6	446	10	BM373545

18	48	6.6	559	10	BE405069
19	48	6.6	588	10	BF255157
20	48	6.6	604	9	AV941745
21	48	6.6	606	10	BF255408
22	48	6.6	631	9	AL505254
23	48	6.6	635	9	AV932837
24	48	6.6	637	9	AV933847
C 25	48	6.6	666	10	BF628036
C 26	48	6.6	719	9	AV935472
27	48	6.6	830	10	BF254391
28	48	6.6	844	12	CNS0052P
29	48	6.6	1101	12	CNS017SY
30	47.8	6.5	307	9	AU070991
31	47.8	6.5	578	9	AU094712
32	47.8	6.5	623	9	AU032635
33	47.6	6.5	517	9	AV933848
34	47.6	6.5	935	12	CNS006XK
35	47.4	6.5	407	10	BE475922
36	47	6.4	657	10	B1717319
37	47	6.4	677	9	AL505169
38	46.8	6.4	269	10	BE636900
C 39	46.8	6.4	641	9	AU093820
C 40	46.8	6.4	669	9	AU096005
C 41	46.8	6.4	674	9	AU055758
42	46.6	6.4	691	9	AL504955
43	46.6	6.4	839	12	CNS004NB
C 44	46.4	6.3	453	10	B1779131
45	46.4	6.3	557	10	BF618612

ALIGNMENTS

RESULT 1
LOCUS CNS0091P/c
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)

REFERENCE

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using the BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoshi Osoegawa and Aaron Hammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"


```

OY      511   tgggaagctcgtatgacgcactggaaattaccatcgccctaacggcgcagaacgcc 570
Db      375   TTCGAGTCTGTGACACCAGATGGGAAGAACAAGCAGCGCAACTTGTCACCATGAGCACGAC 434
OY      571   accacgtcgtgtgagcagacctggaaaggcccttcatacgacgacgcggtcgcccgcgc 630
Db      435   GGCTACTTCCGGACGCGCCCTCGGCTTTCAAAGCCCCCGCAAACTTCACCATCATCTCCGCC 494
OY      631   tacatcgccgcgaagtgtgta 650
Db      495   GACGCCGGGATCTGTCCGTA 514

RESULT       5
CONS03WG/c
DEFINITION   Drosophila melanogaster genome survey sequence TEI3 end Of BAC #
LOCUS        fly), genomic survey sequence.
ACCESSION    AL065254
VERSION      AL065254.1 GI:4942606
KEYWORDS     fruit fly.
SOURCE       Drosophila melanogaster
ORGANISM     Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 964)
AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 Evry cedex - FRANCE (E-mail : seqre@genoscope.cns.fr
              Web : www.genoscope.cns.fr)
COMMENT      determination of this BAC-end and sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mammeter in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              p1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             location/qualifiers
                 source          1..964
                                 /organism="Drosophila melanogaster"
                                 /db_xref="taxon:7227"
                                 /clone_lib="RPCI-98"
                                 /clone="BACR09E09"
                                 /note="end : TEI3"
BASE COUNT           234 a         160 c          60 g          105 t          405 others
ORIGIN
Query Match      7.3%; Score 53.8; DB 12; Length 964;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 91; Conservative 154; Mismatches 180; Indels 1; Gaps 1;
```

```

Db      834  ANMSSSGSGSGSGSGSVSRTRKKRRSSSGBSGGGKSGGGRGAGGGGGGGGGRG  775
Oy      466  ccggcgccgacgcgcgcgttgcacccgttggaactgcccggggcccaactgggaagtcgtcat  525
Db      774  SSKSSTSCGSCSGGRRRTTTRTKGSSSSGSGGTGSGGAAAGSGSGVAGBBGRTGT  715
Oy      526  gccgcgttggaactgaatatacatgcctaccgcggcgacgaagccacacgctcggtgac  585
Db      714  GSTTGSKKGSKYTTTSGKVNNTSSRSYRSTRSGAAROTTTBGSAAAGNSS-  656
Oy      586  gagctggacactgaagacctcatcgacgaacgcggtgcgcgcgcggtacatccgcggag  645
Db      655  SASSVAGVSCCSCMCSASSSCSSSCGCMVCGGCKGCTKCSSSVRSRSGSRTGCGVCGSG  596
Oy      646  tggatctcgactcggttggagacggcgttgaaactctgagggggggggcggtctgcga  705
Db      535  SGBTTTSGGSAAKACACACGCAKGGSSVSGGCGCTYTWSSRADCAAGMACRRVSSGTVA  536
Oy      706  agcgcc 711
Db      535  VRBTSH 530

RESULT 6
LOCUS   AG060149 700 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-047122.R, genomic survey sequence.
ACCESSION AG060149
KEYWORDS  AG060149.1 GI:16611379
SOURCE   GSS; GSS (genome survey sequence).
          Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
          BAC library clone:PTB-047122.R.
          Pan troglodytes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
ORGANISM
          1 (sites)
REFERENCE 1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          2 Tokoki,Y., Watanabe,H. and Sakaki,Y.
          BAC end sequences of library PTB
          Unpublished
          2 (bases 1 to 700)
AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          3 Tokoki,Y., Watanabe,H. and Sakaki,Y.
          Direct Submission
          Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Shohiro-chou,Tsuriumi-Ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
          Clones are derived from the chimpanzee BAC library PTB This BAC end
          was generated during the R&D process and may have higher chance of
          clone tracking errors.
          PRIMERS
          Sequencing: M13Rev
          LIBRARY
          Vector : pKS145
          R.Site 1 : SacI
          R.Site 2 : SacI.
          Location/Qualifiers
          1. 700
             /organism="Pan troglodytes"
             /db_xref="taxon:9598"
             /clone="PTB-047122.R"
             /sex="male"
             /cell_type="lymphoblast"
             /clone_lib="PTB Chimpanzee Male BAC Library"
             /clone_id="PTB 334 g 15 t 30 others

BASE COUNT 30 a 291 c 334 g 15 t 30 others
ORIGIN
Query Match 6.9% Score 50.6 DB 12 Length 700;
Best Local Similarity 42.7% Prid. No. 11;
Matches 243; Conservative 0; Mismatches 323; Indels 3; Gaps 1;

```


LOCUS	AU095988	682 bp	mRNA	linear	EST 30-JUN-2000
DEFINITION	AU095988 Rice green shoot Oryza sativa cDNA clone S11827,				mRNA
ACCESSION	AU095988				
VERSION	AU095988.1	GI:8858670			
KEYWORDS	EST.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa				
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.				
AUTHORS	1 (bases 1 to 682)				
TITLE	Sasaki,T. and Yamamoto,K.				
JOURNAL	Rice cDNA from green shoot (2000)				
COMMENT	Unpublished (2000)				
	Contact: Takuji Sasaki				
	National Institute of Agrobiological Resources				
	Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki				
	305-8602, Japan				
	Tel: 81-298-38-7441				
	Fax: 81-298-38-7468				
	Email: tassaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/PROJECT="RGP".				
	S11827_102.				
FEATURES	Location/Qualifiers				
source	1..682				
	/organism="Oryza sativa"				
	/strain="Nipponbare"				
	/db_xref="taxon:4530"				
	/clone="S11827"				
	/clone_11b="Rice green shoot"				
	/note="Green shoot (8 days old)"				
BASE COUNT	153 a 184 c 212 g 132 t				1 others
ORIGIN					
Query Match	6.8%; Score 49.6; DB 9; Length 682;				
Best Local Similarity	47.1%; Pred. No. 16;				
Matches 185; Conservative	0; Mismatches 205; Indels 3; Gaps 1;				
Oy	248 tcgggtgcacatcgagcgccctcagcagagaattcggaatgcgcgcgacgctgcagagac	307			
Db	453 tggcctcctgcgcagagctggtcttcgaacagaccgagatcctgttcacagctccagta	394			
Oy	308 tctcgaagctgcgcagcagctgcagctcaccgcagatccagcagcgccgtggaatgcc	367			
Db	393 cgcccgacattggtcagcgctcgccgctccaccacacctgcgccacgcgcttcccgagctccg	334			
Oy	368 cctagagacatctgctcagctccgcctcagaaattccgcgaagcagctacagcgcgcgcgcg	427			
Db	333 cctccgacacgaaccccttcttcacgaacaccttggaacgcgcgggaacagaccggaaca	274			
Oy	428 agctgaatcatctgctgaacttgaacgcgcgcgtatgcgcgcgcgcgcgcgcgcgcgcgc	487			
Db	273 ggggagatgctgcgcggaacacagccctgcgtgcgatgcaccccggtgcacgcgacgca	214			
Oy	488 ccgtggaactgc	547			
Db	213 acgtatcctccgctc---cctgtggaattctccggggaacaccttcgacatgctacagatt	157			
Oy	548 tcgacctacgc	607			
Db	156 taccagatcttgctgttccttgcgtacgccttggtggcgccgcgcgaaccttccggccgtgatca	97			
Oy	608 tcgaacgc	640			
Db	96 tgcgcgacccgcttgc	64			
RESULT 13					
A2934730	534 bp	DNA	linear	GSS 24-APR-2001	
LOCUS	A2934730/c				
DEFINITION	BJ_Bs00002f24r B. japonicum BAC library Bradyrhizobium japonicum				

[illegible]

genome
Genome Res. 11 (8), 1434-1440 (2001)
JOURNAL MEDLINE
21376150
COMMENT Contact: Wing RA

```

/note="Vector: pindig0536; Site_1: HindIII"
BASE COUNT      100 a      171 c      190 g      84 t      1 others
ORIGIN

```

Query Match	6.7%	Score 49.2	DB 12	Length 546
Best Local Similarity	51.4%	Pred. No. 19		
Matches 114, Conservative	0	Mismatches 108	Indels 0	Gaps 0

[illegible]

RESULT	15
LOCUS	AU030637/c
DEFINITION	AU030637 Rice cDNA from immature leaf including apical meristem Oryza sativa cDNA clone E50041_27, mRNA sequence.

```

FEATURES
source      Location/Qualifiers
1. 644      /organism="Oryza sativa"

```

BASE COUNT	ORIGIN
150 a	163 c 197 g 127 t 7 others

Query Match	6.7%	Score 49.2	DB 9	Length 644
Best Local Similarity	46.8%	Pred. No. 19		
Matches 184	Conservative	0	Mismatches 206	Indels 3
				Gaps 1

OY	248	tcgtagcgaacttgggagccgcgcgaacgaacattcggaatgcccgcgcgcgtgacgaagc	307
Db	403	TGGCCTCTCGGAGAGCGTGGTTCTCGAAGCAGCGCCGAGTCTTTGTTCACGTCACAGTACA	344
OY	308	tgtagcgaacgtgacgaagaagctgtgaacgctctcaagccgatalcaagacgcgcgcgtgtgaatgcg	367
Db	343	CGCGGAGATTGGTACAGGCTCGGGTGCGCCACACACGCGGCGGCTTTCCCGGACATCCG	28
OY	368	cctaagacaatcttgattcaagttccgctcaacgaattccgcgaacgctctacaagcgcgcgcgcg	427
Db	283	CCTCGGACACGAACCCCTTCGTACGAACCCGCTGAGACGGCGGGAGACAGCAGCCGGACACA	224
OY	428	agctgaatcatctgctgaacactggaagcagcgctgafatgacggcgagcgccgcgtgacga	487
Db	223	GGGGATATTGCTTCGGGGAACAAGCCGTCGTGGGATGTCACCCCGGGGTANAGCCACCGCA	164
OY	488	ccgtggaactctgcgcggcgacacacttgggaagctctgatalgcgcgaacttggatctgaaattaca	547
Db	163	ACGTGATCCCGGTCT---CTCTGTGGGATCTCCGGTGSAACTCTGCAATGCTCAGCATGT	107
OY	548	tcgcctacacgcgcgaacgaacgcccaacacgctctgtgtgaagcagcttgaacctggaagccttca	607
Db	106	TACAGATTTCTGTCCTTTGATACCCCTTGGCGCAGTCGAGACGCTCTCCGCGCCGTGATCA	47
OY	608	tcgaagcaacgcgtgcccgcgcgtcatatccgcg	640
Db	46	TGCGCGACCCGTTCTGCCCGCGGAGCCCGCCGCG	14

Search completed: August 19, 2002, 10:59:56
Job time: 4375 sec

